

# Grau en Estadística

---

**Títol:** Anàlisi de les característiques dels sol·licitants de plaça a estudis universitaris de Barcelona

**Autor:** Judith Arnau Galimany

**Director:** Ernest Pons Fanals

**Departament:** Econometria, Estadística i Economia Aplicada

**Convocatòria:** Gener 2021



UNIVERSITAT DE  
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat de Matemàtiques i Estadística



## **AGRAÏMENTS**

En primer lloc volia agrair a l'Ernest Pons tot el temps que ha dedicat a guiar-me en aquest treball i el recolzament que m'ha proporcionat en tot moment. Sense la seva ajuda no hagués estat possible la realització de l'estudi. Per altra banda, també agrair el suport per part de la meua família i els amics.

## **RESUM**

En aquest projecte es durà a terme a partir d'unes dades obtingudes a través de l'Oficina d'Accés a la Universitat, en concret de l'any 2016. Aquestes dades corresponen als estudiants que han sol·licitat plaça en alguna de les universitats públiques de Catalunya.

L'objectiu del treball es centra en analitzar els estudiants que habiten a Barcelona. S'estudiarà la possible influència d'un seguit de variables, tant socio-demogràfiques com econòmiques, en la decisió de la branca d'estudi de cada individu. Una de les variables que causa més interès és el barri o districte en el que habiten.

A partir d'un seguit d'anàlisi descriptius i de tests estadístics, s'arribarà a una conclusió sobre quins són els factors contemplats que més relació poden tenir en aquest procés.

**Paraules clau:** Barri, districte, variable resposta, variables explicatives, branca d'estudi, renda, correlació, independència.

**Classificació AMS:** 62P25 – Applications of statistics to social sciences



## **ABSTRACT**

This project will be based on a database obtained through the Office of Access to the University. The database is from 2016 and contains information of the students who have presented an application to one of Catalonia's public universities.

The aim of the study is to analyse the students who reside in Barcelona. It will be studied the possible influence of a group of variables to the decision of which field of study they are interested in. Those variables can involve social, economic and demographic information. The most important variable is the neighbourhood or district where they reside.

From a series of descriptive analyses and statistical tests, we will arrive to a conclusion about which are the factors that we have been considered that are more significant in this process.

**Key words:** Neighbourhood, district, outcome variable, covariates, study field, income, causal relationship, independence.

# ÍNDIX

I.	INTRODUCCIÓ.....	1
II.	CONTEXTUALITZACIÓ DE LES BASES DE DADES.....	3
1.	Base de dades principal.....	3
2.	Base de dades codis postal – barris de Barcelona .....	4
3.	Base de dades renda familiar mitjana.....	5
3.1	Definició de RFD.....	5
3.2	Definició de l'indicador .....	6
3.3	Interpretació de l'indicador .....	7
4.	Base de dades codis AQU .....	7
III.	METODOLOGIA .....	8
IV.	TRACTAMENT DE LES BASES DE DADES .....	10
1.	Criteri de divisió codis postal – barris .....	10
2.	Tractament base de dades codis AQU .....	12
3.	Recodificació de variables .....	13
4.	Tractament de <i>missings</i> .....	13
V.	DESCRIPTIVA GENERAL .....	15
1.	Introducció de les variables.....	15
2.	Descriptiva univariant .....	16
VI.	VARIABLES ESCOLLIDES .....	21
1.	Variable resposta .....	21
2.	Variables explicatives.....	22
2.1	Variable Nacionalitat .....	22
2.2	Variable Sexe .....	23
2.3	Variable Edat.....	24
2.4	Variable Renda mitjana .....	25
VII.	ANÀLISI DE LES DADES .....	26
1.	Proporcions d'estudiants universitaris envers índex de renda .....	26
2.	Proporcions dels estudiants envers branques d'estudi .....	27
3.	Relació entre branca d'estudi i gènere .....	29
4.	Relació entre branca d'estudi i nacionalitat.....	31
5.	Relació entre branca d'estudi i Índex de renda familiar mitjana .....	32
6.	Relació entre branca d'estudi i grups d'edat.....	36
VIII.	CONCLUSIONS.....	38
IX.	BIBLIOGRAFIA I WEB GRAFIA.....	40

<b>X. ANNEX.....</b>	<b>42</b>
<b>1. Taules complementàries del registre.....</b>	<b>42</b>
1.1 Valors taula OCB .....	42
1.2 Valors taula O25 .....	42
1.3 Valors taula TM i CE .....	43
1.4 Valors taula ES .....	43
1.5 Valors taula OC .....	43
1.6 Valors taula OR .....	44
1.7 Valors taula CS .....	44
1.8 Valors taula TI .....	44
<b>2. Taula de codificació dels barris.....</b>	<b>45</b>
<b>3. Taula de codificació dels districtes .....</b>	<b>47</b>
<b>4. Taula equivalència dels graus .....</b>	<b>47</b>
<b>5. Taula índex de renda familiar mitjana per barris .....</b>	<b>48</b>
<b>6. Recodificació de les variables .....</b>	<b>50</b>
<b>7. Graus que no apareixien en la classificació .....</b>	<b>52</b>
<b>8. Descriptiva de les variables escollides .....</b>	<b>53</b>
8.1 Variable Nacionalitat .....	53
8.2 Variable Sexe .....	53
<b>9. Taula amb percentatge d'estudiants per barri .....</b>	<b>54</b>
<b>10. Codi utilitzat en R.....</b>	<b>56</b>
10.1 SCRIPT: Depuració del codi postal .....	56
10.2 SCRIPT: Tractament de la base de dades .....	59
10.3 SCRIPT: Descriptiva .....	69
10.4 SCRIPT: Proporcions i anàlisi de dades .....	78

## ÍNDIX DE FIGURES

Figura 1: Separació del districte de l'Eixample per codis postals .....	11
Figura 2: Descriptiva general de les variables categòriques.....	17
Figura 3: Density plot de les variables numèriques contínues .....	18
Figura 4: Boxplot de les variables numèriques contínues .....	19
Figura 5: Gràfic de barres de la variable Ordre d'assignació .....	20
Figura 6: Conjunt de gràfics de barres de les freqüències de cada branca d'estudis .....	21
Figura 7: Mapa per districtes amb pie chart de la variable Nacionalitat.....	22
Figura 8: Mapa per districtes amb pie chart de la variable Sexe.....	23
Figura 9: Mapa per districtes de la mitjana d'edat dels estudiants.....	24
Figura 10: Mapa per districtes de la variable Índex de renda mitjana .....	25
Figura 11: Mapa per districtes de la branca més repetida i el seu percentatge d'estudiants.....	27
Figura 12: Mapa per districtes de la segona branca més repetida i el seu percentatge d'estudiants.....	28
Figura 13: Mapa per districtes de la tercera branca més repetida i el seu percentatge d'estudiants.....	28
Figura 14: Gràfic de barres de les branques d'estudi en funció del gènere .....	30
Figura 15: Gràfic de barres de les branques d'estudi en funció de la nacionalitat.....	32
Figura 16: Conjunt de gràfics de correlació segons branques.....	34
Figura 17: Gràfic de la distribució dels coeficients de Spearman.....	35
Figura 18: Gràfic de barres de les branques d'estudi en funció de grups d'edat .....	37

## ÍNDIX DE TAULES

Taula 1: Recompte de missings en les variables.....	13
Taula 2: Descriptiva numèrica de les variables .....	18
Taula 3: Resultats del test de correlació de Spearman .....	27
Taula 4: Resultats del test d'independència.....	29
Taula 5: Resultats del test d'independència .....	31
Taula 6: Resultats test de correlació de Spearman .....	34
Taula 7: Codificació valors batxillerat-COU .....	42
Taula 8: Codificació valors d'opcions proves per a més de 25 anys .....	43
Taula 9: Codificació valors del grau d'estudis dels pares .....	43
Taula 10: Codificació valors d'ocupació dels pares .....	43
Taula 11: Codificació valors d'orientació.....	44
Taula 12: Codificació dels criteris de selecció .....	44
Taula 13: Codificació de la titulació dels estudiants.....	44
Taula 14: Codificació dels barris.....	46
Taula 15: Codificació dels districtes .....	47
Taula 16: Codificació codis AQU 1r nivell .....	47
Taula 17: Índex de renda familiar mitjana per barri.....	49
Taula 18: Recodificació de les variables .....	52
Taula 19: Percentatges per districtes variable nacionalitat .....	53
Taula 20: Percentatges per districtes variable sexe .....	53
Taula 21: Percentatge d'estudiants per barri .....	55

## I. INTRODUCCIÓ

Una de les decisions més importants en la vida de l'estudiant és decidir quin nivell d'estudis es vol assolir. En el cas que l'estudiant decideixi que vol cursar un grau universitari, es presenta un procés llarg i complex. Quan una persona ha d'escollir la carrera universitària que cursarà es troba en una situació en la que s'ha de plantejar moltes coses que abans no havia necessitat de pensar. Per a la selecció s'ha de buscar informació i analitzar interessos, en general, una exploració interna que no sol ser fàcil.

Es sap que el context en que es troba cada estudiant és un factor el qual pot afectar de manera positiva o negativa. En aquest context influeixen, de manera més generalitzada, dos grans factors. Els interns i els externs. Els interns suposen la part més emocional i personal. En aquest bloc s'hi pot trobar la identificació dels gustos personals, les característiques de cada persona, els interessos, les habilitats i les aptituds, entre d'altres. Aquests factors són els que són més difícils d'analitzar ja que no són comparables entre d'altres. Ja que cada persona és diferent en tots els aspectes i, difícilment, es comparteixen un conjunt de qualitats. Per altra banda, existeixen els factors externs. Acostumen a estar definits per estructures socials, econòmiques i culturals. Aquests factors són el mitjà en el que es desenvolupa i es situa l'estudiant a l'hora d'escollir. En aquest segon bloc es troben variables més demogràfiques que depenen de més gent. Com ara l'àmbit social en el que es viu i la localització entre d'altres.

A l'hora de plantejar el tema d'investigació d'aquest treball, es va voler centrar més en els factors externs. Es coneix que dintre d'aquest grup el factor econòmic té un alt grau d'influència. Partint que les universitats es divideixen en dos grans grups: les públiques i les privades s'ha decidit centrar en un sol grup, les públiques, ja que són més accessibles per a tothom. Un dels altres factors externs que s'ha limitat és la residència dels estudiants. S'ha volgut centrar en els que resideixen a la ciutat de Barcelona.

L'objectiu principal es centra, dintre d'aquests factors externs, en saber si el barri en el que viuen els estudiants té influència a l'hora d'escollir la branca d'estudi del grau universitari. Com a objectius secundaris, interessa poder profunditzar en la resta de factors externs com ara l'econòmic.

Tot i els objectius, l'anàlisi que es durà a terme també permetrà conèixer la distribució dels estudiants universitaris i les seves característiques a la ciutat de Barcelona. És conegut, per estudis anteriors, que el nombre d'estudiants que decideixen accedir a un grau va en augment al llarg dels anys, però també pot ser interessant conèixer si el factor barri pot influir en el nombre d'estudiants que arriben a estudis universitaris.

## II. CONTEXTUALITZACIÓ DE LES BASES DE DADES

En aquest projecte s'utilitzarà principalment quatre bases de dades diferents. A continuació es farà una breu descripció sobre cada una d'elles especificant d'on s'han extret i què inclouen.

### 1. Base de dades principal

En el moment que un estudiant decideix accedir als estudis de grau de les universitats públiques catalanes i la Universitat de Vic-Universitat Central de Catalunya procedeix a formalitzar la preinscripció universitària. Per tal de poder-la realitzar, cada estudiant ha d'omplir un qüestionari. Les dades obtingudes estan gestionades per la Oficina d'Accés a la Universitat. Per a fer possible el tractament, els fitxers contenen les dades individuals fetes anònimes, per tal de preservar la confidencialitat de la informació. Tot i això, per a poder identificar i diferenciar a cada alumne s'ha assignat un identificador a cada un d'ells. Les dades sobre les quals es farà l'estudi són dades recollides sobre els estudiants que cursaran el primer any de carrera el curs 2016-2017.

A continuació es fa una breu descripció del contingut de les dades del qüestionari. El forma un total de 148 variables. Aquestes es poden dividir en vuit blocs:

1. Dades de preinscripció: On recull un seguit de dades personals de l'alumne, com ara el sexe, nacionalitat, codi postal de la residència entre d'altres i informació sobre les proves de la selectivitat, com ara les modalitats escollides, la via d'accés, notes i el centre assignat entre d'altres.
2. Identificació i entorn familiar: En el qual es recull informació sobre els estudis i ocupació dels pares i també l'entorn de cada alumne, com ara si l'alumne treballa i país de naixement.
3. Estudis cursats anteriorment i modalitat d'ingrés a la universitat: Recull dades sobre els estudis anteriors de cada alumne per a conèixer el nivell d'aquests i la modalitat de la qual ha accedit al grau.



4. Idiomes i coneixements informàtics: On s'informa del nivell d'idiomes de cada alumne, el seu grau de coneixement sobre la informàtica i les eines que utilitza.
5. Dades vàries: Ho componen dues variables. Una per a indicar si es té alguna discapacitat o si és esportista d'alt nivell i l'altra per a identificar els becats.
6. Dades matrícula al tancament: Que recull informació sobre el centre en el que serà matriculat i la via d'accés.
7. Bonificació o exempció de taxes: On es tracta si es té algun tractament especial envers a temes econòmics. Per exemple en cas de ser família nombrosa o monoparental i la vigència dels títols.
8. Dades acadèmiques fase general: On s'especifiquen les notes de l'expedient de batxillerat i de la fase general de les PAU i modalitats vàries.

La base de dades tractada ha passat per dues fases de recollida de dades. La primera és quan es fa la recollida de dades general, en la qual existeixen les 148 variables però hi ha que estan en blanc ja que encara no s'ha dut a terme les PAU. La segona fase és quan ja ha passat les PAU i es pot fer el procés d'assignació i matriculació dels estudiants.

Tot i que la base de dades conté moltes variables i estudiants de tot Catalunya, per a la realització d'aquest treball s'ha aplicat un filtre el qual permetrà només agafar els estudiants que visquin a Barcelona, ja que és l'objectiu del treball. La base de dades inicial, amb tots els estudiants, conté 53.949 individus i un cop aplicat el filtre es queda amb 10.154 individus.

## 2. Base de dades codis postal – barris de Barcelona

Aquesta base de dades ha estat generada a partir d'un programa creat amb R el qual ha servit per a poder distribuir tots els estudiants, dels quals es disposava el seu codi postal, en barris de Barcelona.

Aquesta base de dades, anomenada “AlumnesBarris” està composta per dues fulles. La primera està composta per dues columnes, la primera correspon a l’identificador de cada alumne resident a Barcelona i la segona un nombre (entre l’1 i el 75) corresponent cada nombre a un barri de Barcelona.

La segona fulla de l’arxiu conté un llistat de cada barri de Barcelona amb el seu nombre assignat. També conté una altra columna on hi ha un recompte, per barri, del total d’alumnes que conté.

Per a poder assignar cada estudiant, amb el seu respectiu codi postal, a un barri ha calgut un tractament de dades el qual serà explicat en l’apartat de tractament ja que ha suposat un llarg procediment.

### 3. Base de dades renda familiar mitjana

Aquesta base de dades està extreta de la web de l’Ajuntament de Barcelona, a l’apartat que tenen destinat a estadístiques i dades de la ciutat. La font de l’estudi publicat prové de la oficina Municipal de Dades.

La informació que conté és la renda familiar mitjana de Barcelona fragmentada, territorialment, per barris. Aquesta informació està recollida des de l’any 2000 fins l’any 2017. Tot i això, en la base de dades utilitzada per a l’estudi només s’utilitzarà les dades de l’any 2016. Això és degut que els estudiants dels que es disposa la informació són del curs 2016-2017 i l’any 2016 és en el que escullen quin grau voldran cursar.

Per a la recollida de la informació de la renda familiar a Barcelona va caldre crear un indicador específic: l’índex RFD territorial a Barcelona.

#### 3.1 Definició de RFD

La Renda Familiar Disponible per càpita representa la quantitat de renda que disposen les famílies residents per al consum i estalvi una vegada descomptades les

amortitzacions o consum de capital fixe en les explotacions econòmiques familiars, els impostos directes i les quotes satisfetes de la Seguretat Social.

### 3.2 Definició de l'indicador

La definició de l'indicador està extreta de l'estudi *Distribució territorial de la Renda Familiar a Barcelona*<sup>1</sup>.

L'objectiu principal de l'estudi era estimar la renda mitjana o capacitat econòmica dels residents. La creació específica d'aquest indicador va ser deguda a que hi havia una manca d'indicadors fiables i que fossin comparables amb altres d'àmbit territorial.

“L'estimació de l'indicador es realitza a partir de les dades de Renda Familiar Bruta Disponible (RFBD) a preus corrents que difon periòdicament l'Institut d'Estadística de Catalunya per als municipis catalans de més de 5.000 habitants i la resta de capitals comarcals. Aquesta macromagnitud mesura els ingressos de què disposen els residents d'un territori per destinar-los a consum o estalvi. Es calcula com a saldo del compte de renda de les famílies, és a dir, com a diferència entre recursos i usos:

Renda Familiar Bruta Disponible=

- + Sous i salaris bruts
- + Rendes mixtes d'empreses i autònoms
- + Rendes de capital (interessos, dividends i rendes de lloguer)
- + Prestacions socials (Seguretat Social, atur, jubilació i altres transferències de les Administracions Públiques)
- + Transferències i rendes de l'exterior

---

<sup>1</sup> Institut d'Estadística de Catalunya. Renda familiar disponible bruta. Índex. Comarques i Aran, i àmbits. Metodologia [en línia]. Barcelona: IDESCAT, 30/04/2020. Disponible a: <https://www.idescat.cat/pub/?id=aec&n=941&m=m> >

- Cotitzacions Socials
- Impostos directes sobre les famílies“

Per al seu càlcul es va realitzar un estudi sobre les variables en les que es basava i una ponderació específica la qual es pot trobar a l'estudi publicat.

### *3.3 Interpretació de l'indicador*

En aquest cas, aquest indicador, està en base 100. Això vol dir que aquest nombre ens servirà de guia per determinar la renda mitjana dels residents. Per a una més fàcil interpretació en l'estudi estan publicats uns intervals amb una categoria definida:

- Molt alta: més de 159
- Alta: de 126 a 159
- Mitjana-alta: de 100 a 126
- Mitjana-baixa: de 79 a 100
- Baixa: de 63 a 79
- Molt baixa: menys de 63

## **4. Base de dades codis AQU**

Per a possibilitar una classificació, amb un bon criteri, de les titulacions universitàries existents es va recórrer a la que ha creat l'Agència per a la Qualitat del Sistema Universitari de Catalunya (AQU). L'AQU Catalunya és una agència pública que es dedica a l'avaluació, l'acreditació i la certificació de la qualitat en relació a les universitats i els centres d'ensenyament superior de Catalunya. Aquesta agència disposa d'un catàleg de titulacions amb una classificació jeràrquica, on inclou titulacions universitàries, títols superiors, màsters, doctorats i ensenyaments artístics superiors.

En el cas d'aquest estudi, només es centrarà en la classificació que està creada per a les titulacions universitàries. En el catàleg existeixen uns indicadors per a cada nivell d'agregació. Aquestes agregacions estan dividides en tres nivells. El nivell amb el que es treballarà és el primer, que classifica de manera més generalitzada els diversos tipus de branques que existeixen. Tot i així, també ens interessem pel segon nivell, que és més extens.

### III. METODOLOGIA

Per a la realització d'aquest treball s'ha hagut d'implementar els coneixements adquirits al llarg de tot el grau d'estadística, però en concret de les assignatures de "Software estadístic", "Estadística descriptiva", "Mineria de dades" i "Disseny d'experiments".

A partir de les dades seleccionades s'ha volgut estudiar la relació entre unes variables explicatives i una variable resposta. En concret, escollint un cert nombre de variables socials, demogràfiques i econòmiques, relacionar-les i trobar influències entre elles i la variable principal que és la branca de l'estudi que voldrà cursar cada individu. S'ha volgut centrar en els estudiants que resideixen a Barcelona i classificar-los segons el barri ja que és una variable que anteriorment no s'havia estudiat la seva relació amb els estudis.

L'estudi passa per tres fases essencials. Primerament és el processament de les bases de dades. Com es parteix d'un alt nombre de dades i de procedències diferents ja s'ha dedicat un sol apartat per a contextualitzar tota la informació amb la que s'inicia el procés, definir certs paràmetres i exposar com ha estat classificada. Una vegada han estat totes les dades ubicades corresponentment s'inicia el procés de filtratge de dades. En aquest procés el que interessa és, sobretot, separar els estudiants que resideixen a Barcelona dels que no. Seguidament, es fa un extens tractament de les dades, ja que l'objectiu és treballar amb les dades classificades per barris, però no disposem d'aquesta informació de manera directa. Posteriorment caldrà unir tota la informació adquirida en una sola base de dades, recodificar totes les variables i començar a observar quines són les variables que en el següent pas poden interessar per a buscar relacions amb la variable predictora.

En la segona fase es comença a treballar la descriptiva. Primerament es centra en la descriptiva general, per a poder veure quin és el comportament de cada variable i triar quines són les que podrien aportar informació interessant per a fer un estudi més profund i relacionar-les amb la variable "branca d'estudi". A continuació, un cop escollides les variables més interessants, es procedeix a fer un anàlisi bivariant separant de manera geogràfica les variables, concretament segons districtes, i observar si aquesta classificació ens aporta informació útil per a poder conèixer millor el perfil dels estudiants.

En la última fase, un cop ja tinguem triades quines són les variables explicatives i coneixem el seu comportament, es procedirà a fer un anàlisi tant descriptiu com amb tests estadístics per a poder analitzar les suposicions que s'han anat fent durant tot el transcurs de l'estudi sobre la possible relació o influència de certes variables sobre la variable resposta. En aquesta part del treball es contemplaran varis tipus de gràfics que poden ser molt útils per a veure com actua la variable principal quan es classifica segons les altres variables o també que es pugui veure la possible relació entre elles. També s'utilitzaran tests com ara el test d'independència de la chi-quadrat de *Pearson*.

Un cop acabades totes les fases esmentades es podrà arribar a unes certes conclusions que ens diran si les suposicions inicials es poden acceptar o rebutjar.

## IV. TRACTAMENT DE LES BASES DE DADES

L'objectiu d'aquest apartat és exposar com s'han hagut de tractar i/o generar bases de dades que després seran utilitzades durant el procés d'anàlisi de les dades. Així doncs, s'explicarà l'origen d'aquestes, quin ha estat el procés de tractament i l'objectiu d'obtenir-les.

### 1. Criteri de divisió codis postal – barris

L'objectiu principal d'aquest treball és poder avaluar si influeix o no el barri en el que viu cada estudiant. Una de les condicions per a que la base de dades tractada pugui ser distribuïda és respectar l'anonimat dels estudiants que la componen. És degut a això que la única informació de caràcter més personal que conté la base de dades és el codi postal on viuen. A partir d'aquesta informació disponible es va haver de fer un plantejament sobre com es podia passar del codi postal a distribuir els estudiants en els seus barris corresponents.

La idea inicial era separar cada codi postal en el seu barri corresponent de manera exacta, però aquesta informació no està disponible sense cap càrrec econòmic. A partir d'això, la següent idea va ser separar manualment cada codi postal en el seu respectiu barri. El problema que es presenta és que cada codi postal, la majoria de vegades, pertany a diversos barris. Així doncs, el procés de separació és aproximat i amb un marge d'error que s'anirà acumulant al llarg del treball.

Aquest procés doncs, constarà de quatre parts principals. El primer pas que vaig seguir per a poder-ho fer va ser buscar un mapa de Barcelona dividit per barris. Per a que la separació dels barris fos el més visible possible vaig decidir optar per utilitzar un mapa per cada districte de Barcelona (10). Seguidament, vaig anar marcant manualment la delimitació de cada codi postal guiant-me amb un mapa dels codis postals de Barcelona.

Com que la repartició entre els codis postal i els barris no és exacta, en la segona part del procediment, s'ha hagut de trobar la manera de fer una partició justa per repartir les persones de cada CP amb el seu respectiu barri. Així doncs, per cada barri s'ha observat quants codis postal el componien i s'ha fet una divisió de cada CP en sectors

aproximats. Per exemple, un codi postal pot correspondre un terç a un barri i dos terços a un altre barri. S'ha vist que el màxim de codis postal que coincideixen en un mateix barri són set.

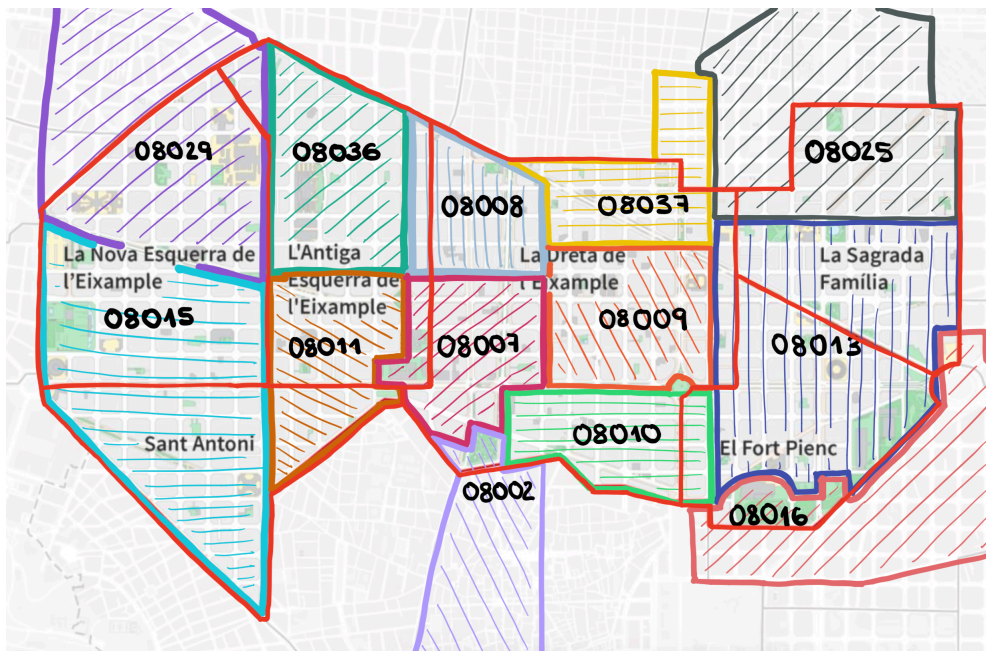


Figura 1: Separació del districte de l'Eixample per codis postals

En tercer lloc, un cop feta la divisió, s'ha passat tota aquesta informació en un Excel per a poder-la tractar amb el programa R. Aquesta base de dades estarà composta per dues fulles. La primera està composta per 16 columnes. La primera conté un llistat de tots els possibles CP existents a Barcelona. La segona, anomenada "divisions" conté el nombre de divisions en que s'ha dividit cada CP. I finalment, de la tercera columna a la última, que són set parelles, contenen el barri corresponent i la porció de CP que li pertoca. En la segona fulla hi ha un llistat on s'especifica quin nombre s'ha assignat a cada barri de Barcelona per a fer el tractament durant la codificació més fàcil. Aquesta base de dades té aquest format perquè és la manera més fàcil que es va trobar per a poder assignar els alumnes a cada barri.

En quart lloc, ve el procés de codificació i tractar la base de dades. La idea principal del codi era que coneixent cada codi postal les seves divisions i proporcions corresponents, pogués recórrer, amb un bucle, cada un d'ells. En aquest bucle, el tractament de les dades correspondria a conèixer el nombre d'alumnes que conté cada CP i dividir-los de manera aleatòria entre les divisions corresponents. I un cop obtingudes les divisions assignar-los, segons les porcions, al barri corresponent. En els casos en que, al fer les divisions, el nombre d'alumnes no fos exacte, s'assignava de manera aleatòria als barris



els alumnes restants. D'aquesta manera, treballant amb l'atzar, s'aconsegueix que la repartició no estigui condicionada per cap factor i així no pugui causar futurs problemes. Tot i així, al llarg de l'anàlisi es tindrà en compte que aquest fet pot arrastrar un cert grau d'error.

Un cop executat el codi per a repartir els alumnes entre els seus futurs barris assignats, s'aconsegueix la base de dades, abans descrita, anomenada "AlumnesBarris". Amb aquest procés s'aconsegueix poder encarar el treball cap a l'objectiu principal d'observar els alumnes segons els seus barris.

## 2. Tractament base de dades codis AQU

A Catalunya actualment existeix una oferta de 471 graus possibles a escollir entre les universitats públiques. En aquest nombre també es contemplen dobles titulacions i triples. Com que seria molt complicat poder comprovar si el barri influeix en quin grau escull cada alumne, s'ha optat per una classificació d'aquests. Així doncs, cal trobar la manera de classificar els graus i aquesta, com s'ha esmentat abans, serà la que publica l'AQU. Per a poder identificar els graus universitaris disponibles s'ha utilitzat els codis que assigna a cada grau la Oficina d'Accés a la Universitat. El tractament d'aquesta base de dades consta de dues parts principals:

La primera correspon a l'assignació manual dels grups. Com que el codi de cada grau que assigna l'AQU no correspon amb el que assigna la OAU caldrà anar un per un dels graus i assignar un codi amb la classificació de l'AQU. Aquesta classificació conté dos nivells. El primer nivell és de caire més general i conté 17 grups mentre que el segon és més concret i conté 46 grups. La idea principal és tractar la influència amb el segon nivell però si el resultat és que no influeix, s'intentarà comprovar amb el primer nivell.

Per altra banda, en la segona part del tractament, existeixen alguns graus en la llista de la Oficina d'Accés que no estan en la llista dels codis AQU. Per tant, caldria trobar una manera per a poder-los classificar. Són un total de 29 graus. Com que és un nombre prou elevat i no ens interessa ometre informació s'ha decidit fer la classificació<sup>2</sup> segons el

---

<sup>2</sup> Adjunt un llistat amb els graus a l'apartat 7 de l'annex.

criteri personal. Tot i això, hi ha graus que donen certs problemes ja que no hi ha classificació que els representi correctament.

### 3. Recodificació de variables

Per poder treballar adequadament amb les dades, en primer lloc s'ha afegit les noves variables que no apareixien en la base de dades original, com ara la variable barri, renda i les classificacions de les branques universitàries.

Quan anomenem de les noves variables afegides, en la base de dades apareixen sis noves variables. Aquestes són PREFE1\_1, PREFE1\_2, CENTRE.ASSIG\_1, CENTRE.ASSIG\_2, CENTRE.MATRIC\_1 i CENTRE.MATRIC\_2. Aquestes noves variables apareixen perquè la seva variable base (PREFE1, CENTRE.ASSIG i CENTRE.MATRIC) corresponen als estudis universitaris, els quals ens interessa tenir-los classificats per les branques que estableix l'AQU.

A continuació s'ha estudiat la tipologia de cada variable i s'ha definit corresponentment com a factor o numèric per a la seva correcta lectura. Per a facilitar l'estudi, també s'ha recodificat diverses variables categòriques i en alguns casos de variables numèriques s'han generat uns intervals englobant un conjunt de respostes.

### 4. Tractament de *missings*

El recompte de *missings* de la base de dades és el següent:

*Taula 1: Recompte de missings en les variables*

Variable	Nombre de <i>missings</i>	Variable	Nombre de <i>missings</i>
ID_ALUMNE	0	ESTU_MARE	0
NACIONALITAT	0	OCUP_PARE	0
CP	0	OCUP_MARE	0
SEXE	0	TREB_REM	0
edatT	0	ORIENTACIO	0
DATA_NAIX	0	ELEC_EST	0
VIA.ACCES	0	TIP_CENTRE	772

OPC.CO.U.BATXILLERAT.MG25	0	TIP_CENTRE_FP	8415
ANY.QUAL	0	EST_ANT	0
NOTA.ACCES.PAU.EXPEDIENT.CFGS	0	TITULACIO	7822
NOTA_ADMISSIO	0	CENTRE.MATRIC	0
SIST_EDU_ES	2602	BARRI	0
PREFE1	0	renda16	0
ULT_CARR_COM	8717	PREFE1_1	498
ULT_CURS_FIN	9196	PREFE1_2	498
CENTRE.ASSIG	0	CENTRE.ASSIG_1	448
ORDRE.ASSIG	0	CENTRE.ASSIG_2	448
CONCOR	10	CENTRE.MATRIC_1	351
ESTU_PARE	0	CENTRE.MATRIC_2	351

Una de les coses a destacar d'aquesta informació és l'aparició de *missings* en les variables creades (PREFE1, CENTRE.ASSIG i CENTRE.MATRIC) ja que en les seves variables originals no se'n detecta cap. Això ve degut ja que d'un total de 723 graus que s'ofereixen, 97 són doble grau o agrupacions de titulacions. Aquestes 97 titulacions engloben un total de 498 alumnes que han sol·licitat plaça en aquests.

A l'hora de tractar aquests alumnes, es van valorar dues idees. La primera era duplicar cada alumne pel nombre de titulacions que agrupés la oferta de grau. Però en aquesta situació s'estava afegint també el doble de la informació d'altres variables, que després a l'hora de realitzar l'anàlisi podia influir notablement. Per altra banda, la opció també era descartar aquests alumnes. Suposa la pèrdua d'un 4,9% d'informació, però és la opció més viable en aquest cas. A l'hora de realitzar l'anàlisi ja es tindrà en compte aquest error acumulat. Amb la resta de variables que tenen *missings* s'ha decidit no incloure-les dins les variables escollides. Això és degut a que no són uns valors que es puguin generar ja que tenen un motiu en concret per no existir. Per exemple, en la variable ULT\_CARR\_COM, hi ha estudiants que no tenen cap carrera començada. En el cas d'escollir algun mètode per iterar aquests *missings* el que s'estaria fent és afegir dades que no poden ser possibles.

## V. DESCRIPTIVA GENERAL

Per a poder conèixer el perfil dels estudiants que formen part de la base de dades que s'està estudiant es realitzarà una descriptiva general de totes les variables. D'aquesta manera també ajudarà a escollir les variables que després formaran part de l'anàlisi.

### 1. Introducció de les variables

A continuació es defineix de manera concisa les variables explicatives que s'han considerat clau entre les 148 variables totals. D'entre aquestes s'escolliran les que es creguin adients per a fer un estudi amb més profunditat.

#### **Categòriques:**

ID\_ALUMNE: Identificador per a cada alumne.

NACIONALITAT: Nacionalitat de cada estudiant.

CP: Codi postal del domicili familiar o residència habitual.

SEXE: Sexe de cada persona.

VIA ACCES: Mètode utilitzat per accedir a la convocatòria d'accés a la universitat.

OPC COU/BATXILLERAT/MG25: Opcions Batxillerat/COU. Segons valors de la taula OCB. Opcions més grans de 25 anys. Segons valors de la taula O25.

ANY QUAL: Any de qualificació de les PAU.

SIST\_EDU\_ES: País amb sistema educatiu estranger sense PAU.

PREFE1: Codi de la preferència segons el llistat vigent per cada convocatòria.

ULT\_CARR\_COM: Última carrera començada. Segons valors de la taula TM.

ULT\_CURS\_FIN: Any de la última carrera finalitzada. Segons valors de la taula TM.

CENTRE ASSIG: Centre assignat. Segons valor de la taula CE.

CONCOR: Concordança entre preferència 1 i centre assignat.

ESTU\_PARE: Estudis acabats del pare. Segons valors de la taula ES.

ESTU\_MARE: Estudis acabats de la mare. Segons valors de la taula ES.

OCUP\_PARE: Ocupació del pare. Segons valors de la taula OC.

OCUP\_MARE: Ocupació de la mare. Segons valors de la taula OC.

TREB\_REM: Treball remunerat de l'alumne.

ORIENTACIO: Orientació a escollir els estudis universitaris. Segons valors de la taula OR.

ELEC\_EST: Aspectes que s'han tingut en compte en l'elecció d'estudi. Segons valors de la taula CS.

TIP\_CENTRE: Tipus de centre on s'ha estudiat l'últim any de secundària.

TIP\_CEN\_FP: Tipus de centre on s'ha estudiat l'últim any de CFGS o FP2.

EST\_ANT: S'ha cursat estudis universitaris anteriors.

TITULACIO: Titulació o nivell assolit. Segons valor de la taula TI.

CENTRE\_MATRIC: Codi del centre matriculat. Segons valors de la taula CE.

BARRI: Barri assignat a cada alumne segons codi postal.

PREFE1\_1: Recodificació de la variable PREFE1 amb codis AQU del 1r nivell.

PREFE1\_2: Recodificació de la variable PREFE1 amb codis AQU del 2n nivell.

CENTRE\_ASSIG\_1: Recodificació de la variable CENTRE\_ASSIG amb codis AQU del 1r nivell.

CENTRE\_ASSIG\_2: Recodificació de la variable CENTRE\_ASSIG amb codis AQU del 2n nivell.

CENTRE\_MATRIC\_1: Recodificació de la variable CENTRE\_MATRIC amb codis AQU del 1r nivell.

CENTRE\_MATRIC\_2: Recodificació de la variable CENTRE\_MATRIC amb codis AQU del 2n nivell.

### **Numèriques:**

edatT: Edat teòrica.

NOTA\_ACCES\_PAU\_EXPEDIENT\_CFGS: Nota d'accés a les PAU / Nota expedient CFGS.

NOTA\_ADMISSION: Nota d'admissió a la universitat.

ORDRE\_ASSIG: Ordre d'assignació del centre d'estudis.

renda16: Renda mitjana familiar per barri de l'any 2016.

## **2. Descriptiva univariant**

Primer s'estudiarà les **variables categòriques**. Es decideix utilitzar els gràfics de barres ja que en alguns casos hi ha variables que presenten un alt nombre de possibles respostes.

## DESCRIPTIVA UNIVARIANT DE LES VARIABLES CATEGÒRIQUES

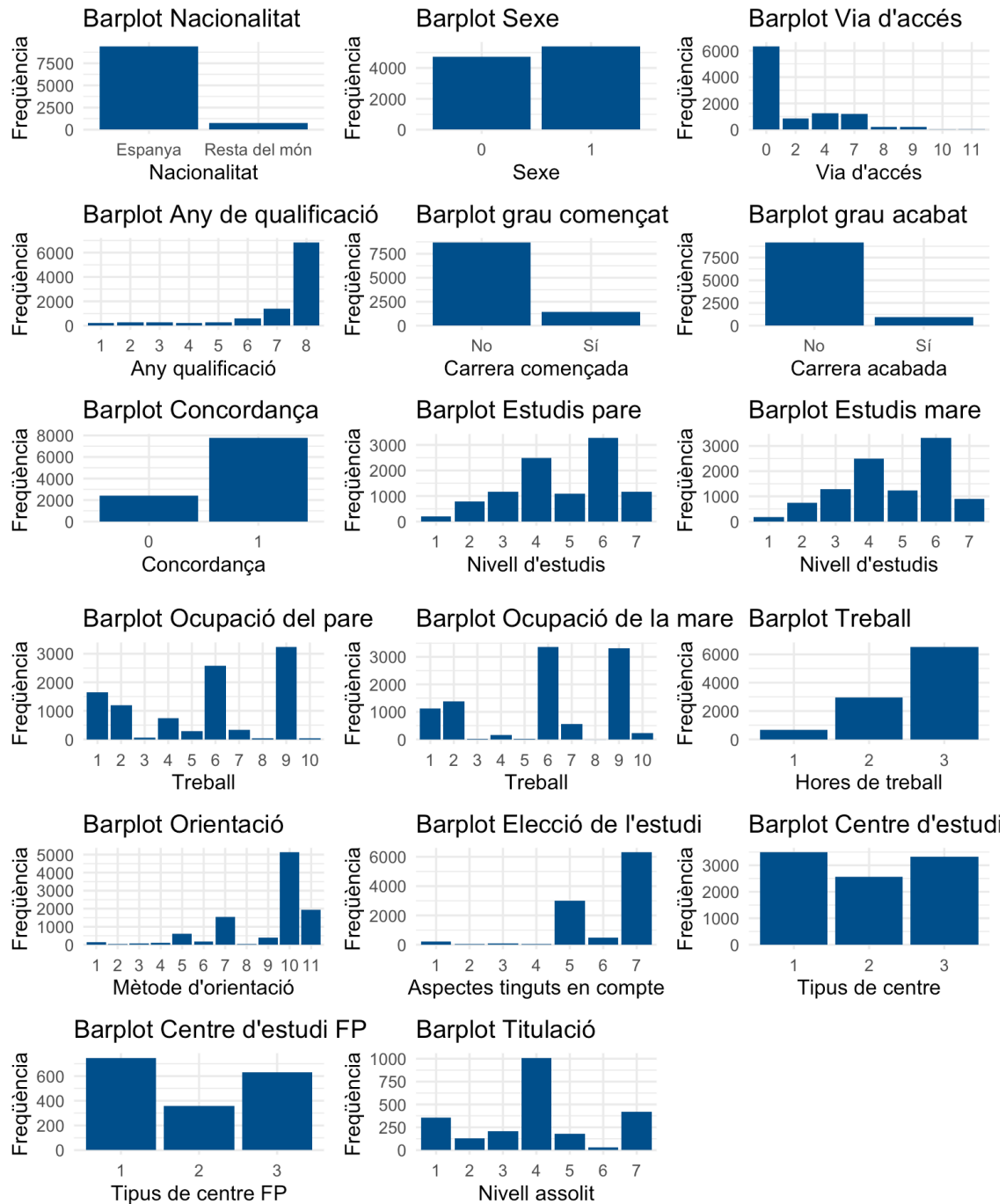


Figura 2: Descriptiva general de les variables categòriques

Hi ha diverses dades a destacar, començant per les variables que estan balancejades, com ara la variable “sexe”, “centre d’estudi” i “centre d’estudis FP”.

Per altra banda, tenim una sèrie de variables en que hi ha un tipus de resposta molt dominant per sobre les altres. Per exemple, en la variable “nacionalitat” es pot veure que hi ha una gran quantitat de persones que són de nacionalitat espanyola envers a la resta.

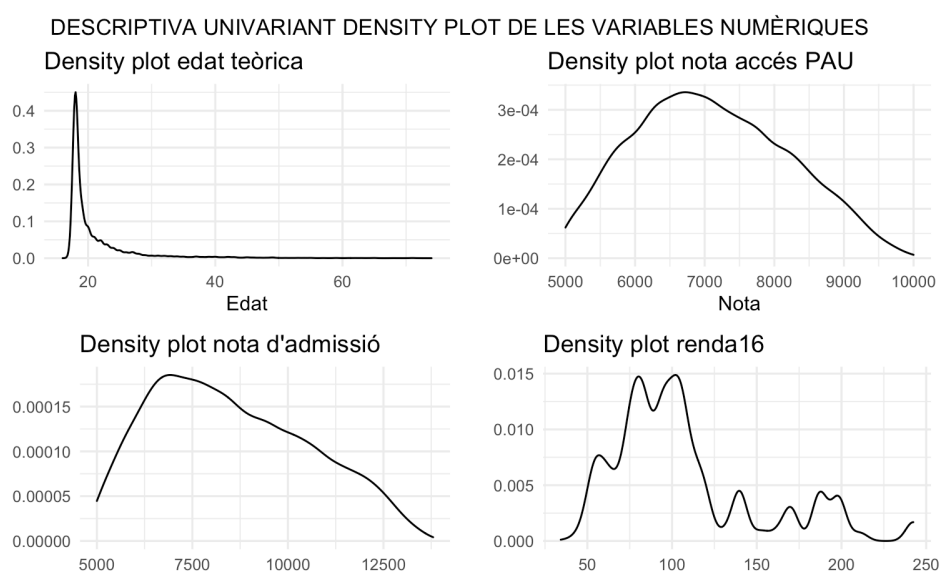
Es veu també en la variable “via d’accés”, on la resposta dominant és la gent que prové de les PAU, la variable “any de qualificació” on la resposta dominant és l’any 2016. Dins d’aquest grup de variables amb respostes molt diferenciades també es pot trobar la variable “grau començat” ja que la gran majoria ha respost que no, i en conseqüència, la variable “grau acabat” també té com a resposta dominant el no. Finalment, destacar la variable “titulació” on els alumnes que havien cursat anteriorment algun grau universitari, la majoria tenien menys de 180 crèdits superats.

Pel que fa al conjunt de **variables numèriques** es realitza una taula amb diversos valors per observar com es distribueixen les respostes.

*Taula 2: Descriptiva numèrica de les variables*

	MIN	Q1	MEDIANA	MITJANA	Q3	MAX
<b>Edat</b>	16,0	18,0	19,0	21,378903	22,0	74,0
<b>Nota accés expedient</b>	5000,0	6270,0	7028,0	7102,97	7896,0	10000,0
<b>Nota d'admissió</b>	5000,0	6872,0	8272,0	8528,01	1006,0	13800,0
<b>Ordre assignació</b>	0,0	1,0	1,0	1,24	1,0	8,0
<b>Renda16</b>	34,3	77,9	95,9	107,37	119,9	242,4

També es realitzarà un *density plot* per estudiar-ne la seva distribució.



Es pot observar que en les variables que contempla notes, en la variable “nota accés PAU/expedient” i en la “nota d’admissió”, segueixen una certa normalitat. També es pot observar que en ambdós casos el pic es centra al voltant del 7 tot i tenir màxims diferents. Per altra banda, es pot veure que la variable “edat teòrica” i “renda16” presenten una distribució bastant irregular amb alguns *outliers* (tots a la cua dreta). En concret, en la variable “edat teòrica” es pot veure que la majoria de individus es centren al voltant dels 19 anys, com mostra la mediana, però existeixen un seguit d’*outliers* a la cua esquerra.

Per a comprovar l’existència d’*outliers* es realitzarà un seguit de *boxplot* per a fer-ho més visuals:

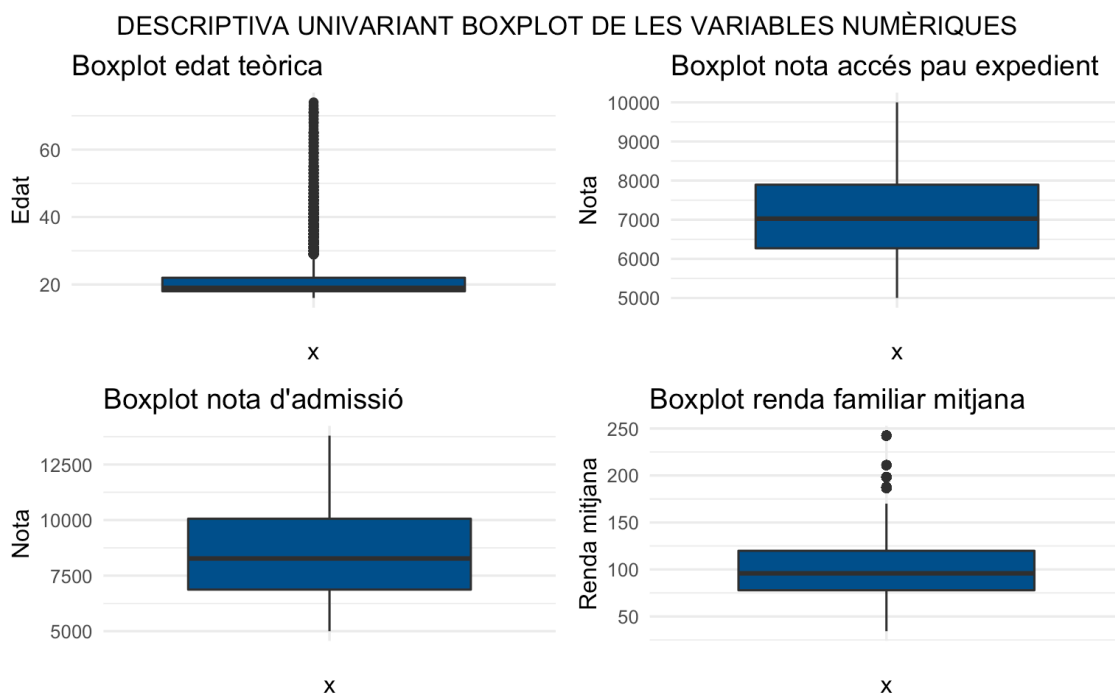


Figura 4: Boxplot de les variables numèriques contínues

Efectivament, com es veia anteriorment, es pot destacar la presència d’*outliers* en la variable “renda16”. Es pot destacar la semblança en la variabilitat entre la variable “nota accés pau expedient” i la variable “nota d’admissió” tot i la seva diferència en l’eix d’ordenades. També en la variable “edat teòrica” podem veure que existeix poca variància i al voltant del 20, i a continuació, un seguit d’*outliers*.

A continuació es gràfica la variable discreta “ordre d’assignació”, la qual ens informa que la majoria dels estudiants entren a la primera assignació i que molt pocs entren a la vuitena. Per altra banda tenim la aparició del zero. Aquests són els que no han estat assignats enlloc.



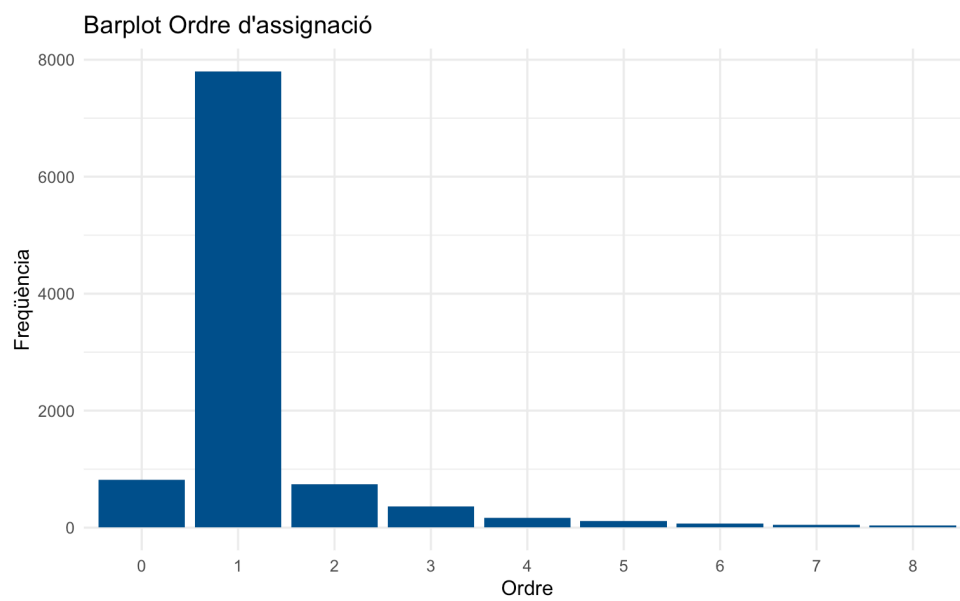


Figura 5: Gràfic de barres de la variable Ordre d'assignació

## VI. VARIABLES ESCOLLIDES

Una vegada es coneixen les variables, és el moment d'escollir quines seran les que s'utilitzaran. Per al procés d'elecció de les variables s'ha fet una revisió de la descriptiva obtinguda anteriorment. També s'han tingut en compte valoracions de variables que semblaven rellevants d'altres estudis previs sobre aquest sector.

### 1. Variable resposta

A l'hora de triar la variable resposta hi havia diverses opcions: *PREFE1*, *CENTRE.ASSIG* i *CENTRE.MATRIC*. Com que el que es pretén estudiar és quina branca d'estudis escull cada estudiant depenent de diverses variables influents, no ens interessa saber ni quin centre se'ls assigna ni on s'acaben matriculant. Per tant, la variable que s'escull com a resposta és la *PREFE1* ja que és la seva preferència inicial. Aquesta variable ha estat recodificada amb els codis corresponents segons l'AQU, així doncs, serà la variable *PREFE1\_1*.

Per a conèixer millor aquesta variable es graficarà la seva informació per districtes.

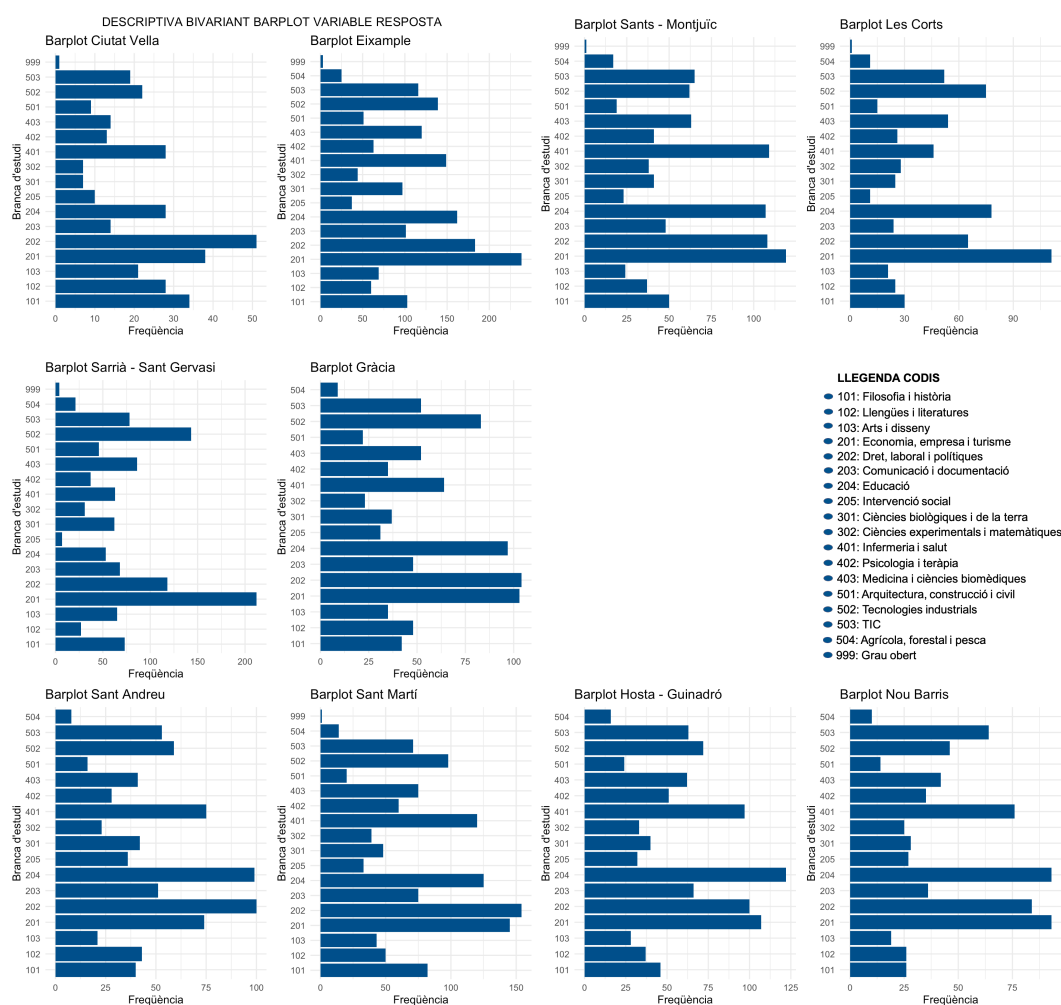


Figura 6: Conjunt de gràfics de barres de les freqüències de cada branca d'estudis

Es pot observar que hi ha certs valors que tenen una alta freqüència en la gran majoria dels districtes, com ara les categories 201, 202 i 204. Però la resta de categories varien molt depenent de quin lloc s'estigui observant. És per això que serà interessant fer un posterior anàlisi junt amb les proporcions dels estudiants de cada branca per a poder veure si el lloc d'on provenen és independent de l'elecció de l'estudi.

## 2. Variables explicatives

Com a variables predictores s'han escollit diversos tipus d'informació. Informació més general, com ara la nacionalitat, el sexe i l'edat teòrica. I finalment, unes variables més concretes, com són el barri on viu i l'índex de renda familiar mitjana del barri on viu.

Per conèixer més les variables es realitzarà una descriptiva bivariant<sup>3</sup> d'aquestes respecte la seva distribució per districtes:

### 2.1 Variable Nacionalitat

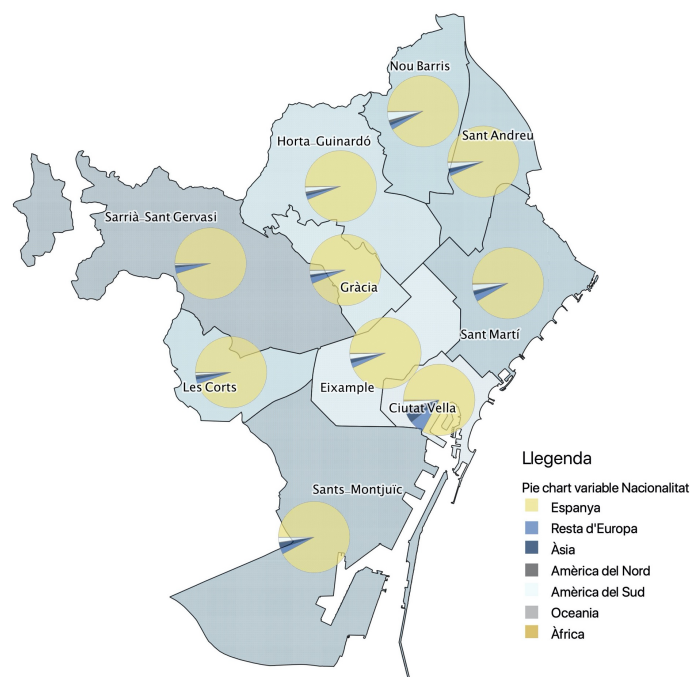


Figura 7: Mapa per districtes amb pie chart de la variable Nacionalitat

El que interessa destacar d'aquesta projecció és, sobretot, l'alta quantitat de persones procedents d'Espanya que es presenten. Tot i això, podríem destacar que en certs districtes, com ara Ciutat Vella i Nou Barris són dels que agrupen més persones d'altres

<sup>3</sup> S'adjunta un seguit de taules amb informació complementària sobre les variables categòriques a l'apartat 8 de l'annex.

nacionalitats. Per un altre costat, també és interessant veure que de les nacionalitats que provenen més persones, en general, són de la resta d'Europa, d'Àsia i d'Amèrica del sud.

## 2.2 Variable Sexe

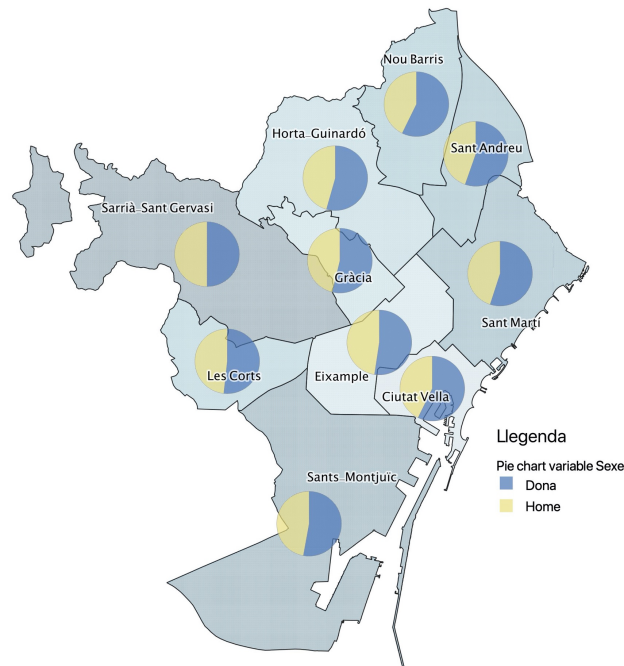


Figura 8: Mapa per districtes amb pie chart de la variable Sexe

En el cas de la variable sexe es pot observar que en tots els districtes les dones estan més presents que els homes, excepte a Sarrià-Sant Gervasi on són equitatius. També caldria comparar amb la proporció d'homes i dones que habiten en cada barri per a poder observar si hi ha alguna informació interessant a destacar.

### 2.3 Variable Edat

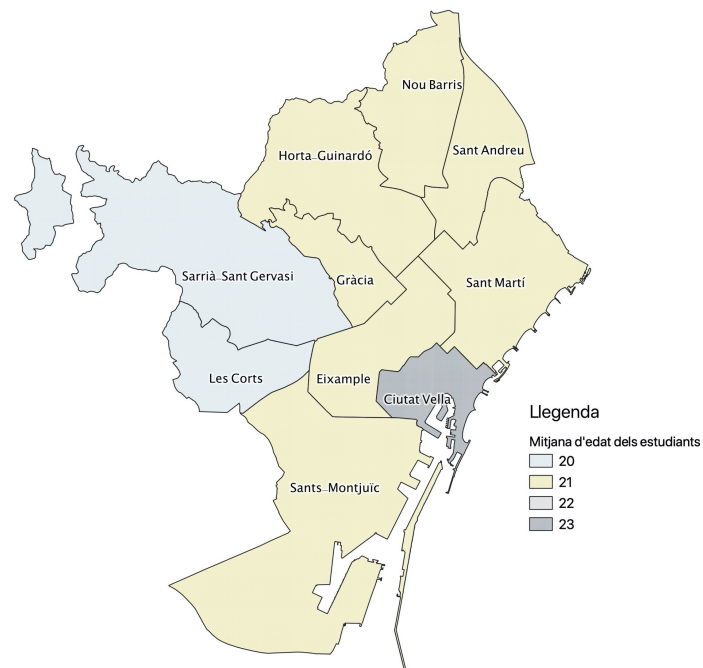


Figura 9: Mapa per districtes de la mitjana d'edat dels estudiants

En aquesta figura es pot observar que la majoria de districtes tendeixen a tenir una mitjana d'edat de 21 anys dels estudiants que es presenten a les proves d'accés. Per altra banda, destaquen dos districtes que la tenen un any més baixa i un que la té més alta. Podria ser una variable interessant a estudiar per veure si l'edat té relació amb el districte o barri on es troben.

## 2.4 Variable Renda mitjana

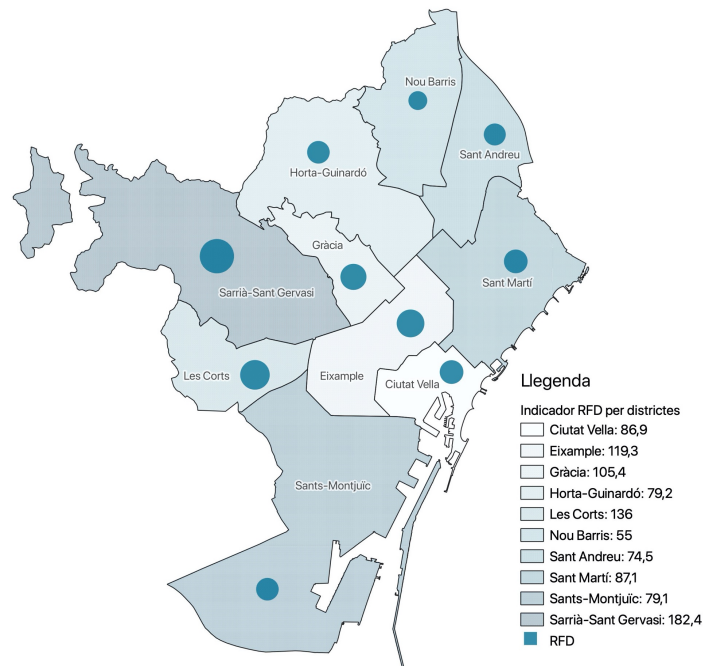


Figura 10: Mapa per districtes de la variable Índex de renda mitjana

En aquesta última figura es pot destacar l'anàlisi de la variable RFD, que correspon a un indicador de la mitjana de la renda familiar segons districtes. En aquest indicador, com més alt és el nombre, més poder adquisitiu tenen les famílies. Consta d'un gràfic representat per proporcions. Com més gran és la circumferència que es troba en el centre de cada districte, més alt és l'indicador. D'aquesta variable interessa saber si té influència en la variable que mostra la branca d'estudi.

## VII. ANÀLISI DE LES DADES

En aquest últim apartat del projecte es procedeix a realitzar un anàlisi de les dades que s'han escollit com a més interessants durant els processos anteriors. A continuació, a partir d'un seguit de tests estadístics i gràfics es comprovarà la hipòtesi inicial, que era saber si el barri influïa en l'elecció de la branca d'estudi dels universitaris. També es comprovaran altres possibles relacions treballant amb les variables seleccionades restants.

Per a treballar amb les dades proporcionades s'ha decidit que el millor és treballar durant l'anàlisi amb proporcions, ja que així facilitarà la comparació. En cada apartat es definirà totes les proporcions utilitzades.

Tot i els resultats obtinguts en els propers anàlisi, cal tenir en compte que les dades tractades només corresponen a l'any 2016, per tant, això afegeix un biaix important.

### 1. Proporcions d'estudiants universitaris envers índex de renda

Per començar aquest anàlisi, una de les preguntes és si la renda familiar mitjana influeix a l'hora d'accedir a estudis universitaris. És conegut que la renda no es distribueix per igual segons el barri on es viu. Per tant, utilitzant el percentatge d'estudiants<sup>4</sup> que pertany a cada barri junt amb l'Índex de renda es realitzarà un test de correlació de *Spearman*.

Amb aquest test el que es pretén comprovar és si existeix una relació o tendència significativa entre les dades, per tant si la correlació és significativament diferent de zero. Un cop comprovat si hi ha relació o no, es tindrà ja una idea inicial sobre què tenir en compte a l'hora de parlar de la distribució d'estudiants segons barris i fer futurs anàlisis. Per tant procedim a realitzar el test a través de codificació d'R.

Primerament cal plantejar les hipòtesis del test de correlació:

$H_0: \rho = 0$  (no existeix relació)

$H_1: \rho \neq 0$  (sí existeix relació)

---

<sup>4</sup> S'adjunta taula amb el recull dels percentatges a l'apartat 9 de l'annex.

Mitjançant la funció 'cor.test(x,y,method = "spearman")' obtenim els següents resultats.

*Taula 3: Resultats del test de correlació de Spearman*

Estadístic	P-valor	Rho (coeficient de correlació)
30.177	8,976e-08	0,5707437

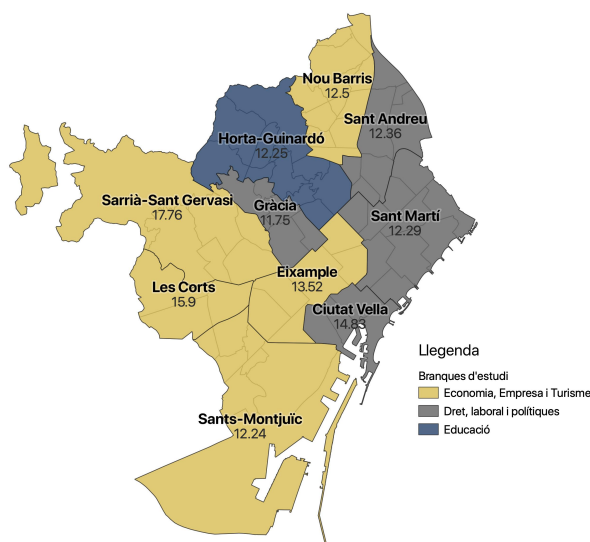
Per a la interpretació del valor *rho* cal saber:

- Al ser un nombre positiu, la correlació que hi ha és creixent, per tant, quan un nombre augmenta, l'altre també.
- Com més proper a l'1 o -1 més correlació hi haurà, per tant ens trobem que hi ha correlació però no molta.

Un cop sabem això, podem procedir amb l'anàlisi i començar a observar les variables segons les branques.

## 2. Proporcions dels estudiants envers branques d'estudi

En aquest apartat es centra en mostrar quines són les branques que obtenen més proporcions d'estudiants segons districtes. A continuació s'exposa en forma de mapa les branques que han aparegut més per cada districte junt amb el seu percentatge.



*Figura 11: Mapa per districtes de la branca més repetida i el seu percentatge d'estudiants*



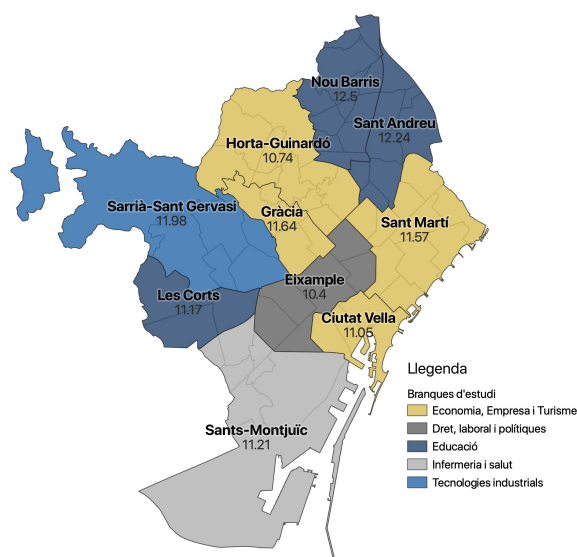


Figura 12: Mapa per districtes de la segona branca més repetida i el seu percentatge d'estudiants

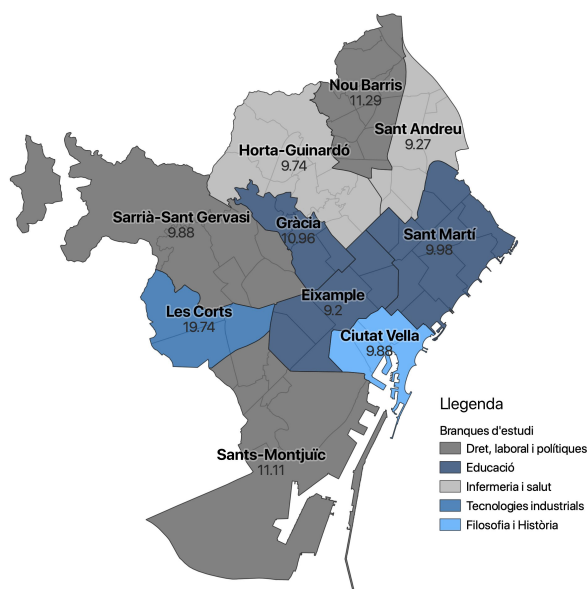


Figura 13: Mapa per districtes de la tercera branca més repetida i el seu percentatge d'estudiants

El que es pot destacar d'aquestes tres figures és que ressaltava molt quines són les branques més escollides entre els sol·licitants. També és interessant veure que hi ha algun districte que es desplaça de la normalitat, com ara Sant Andreu, que és l'únic on no apareix la branca d'economia, empresa i turisme com a una de les tres més comunes. Per altra banda tenim Ciutat Vella, on en el tercer lloc apareix filosofia i història, una branca menys comú.

A continuació es procedirà a analitzar si el barri és independent o no de l'elecció de cada branca. Per a poder-ho fer s'utilitzarà el test d'independència de la  $\chi^2$  de *Pearson* i es treballarà amb les freqüències d'estudiants per cada barri i branca.

Aquest test s'utilitza per a estudiar si existeix o no una associació entre dues variables categòriques. En aquest cas les variables són "Barri" i "Branca". Per a la comprovació d'aquest test treballarem amb una *alpha* del 0,05. Per tant, la hipòtesi del test serà:

$H_0$ : Les variables són independents, per tant, la categoria de la variable "branca" no varia entre els diferents nivells de la variable "barri".

$H_1$ : Les variables són dependents, per tant, la categoria d'una variable pot influir entre els diferents nivells de l'altra variable.

L'estadístic es construeix de la següent manera:

$$\chi^2 = \sum_{ij} \frac{(\text{observats}_{ij} - \text{esperats}_{ij})^2}{\text{esperats}_{ij}}$$

S'obtenen els següents resultats:

*Taula 4: Resultats del test d'independència*

Estadístic	Graus de llibertat	P-valor
1.616,8	1.241	2,492e-12

S'observa un *p-valor* molt més inferior a 0,05. Per tant, tenim suficient evidència per rebutjar la hipòtesi nul·la. És per això que podem dir que l'elecció del grau universitari estarà relacionat amb el barri d'on provinguí cada estudiant.

### 3. Relació entre branca d'estudi i gènere

Un cop estudiada i comprovada la relació entre el barri i la branca d'estudi també interessa poder observar si hi ha altres variables que podrien influir en l'elecció. Per tant, aquest apartat es centrarà en analitzar la variable gènere en funció de les diverses branques.

Per començar, es farà un gràfic de barres de les freqüències distingint entre els dos sexes, home i dona. Cal destacar que la realització d'aquest gràfic per a comparar és possible ja que el nombre d'homes i el nombre de dones és similar.

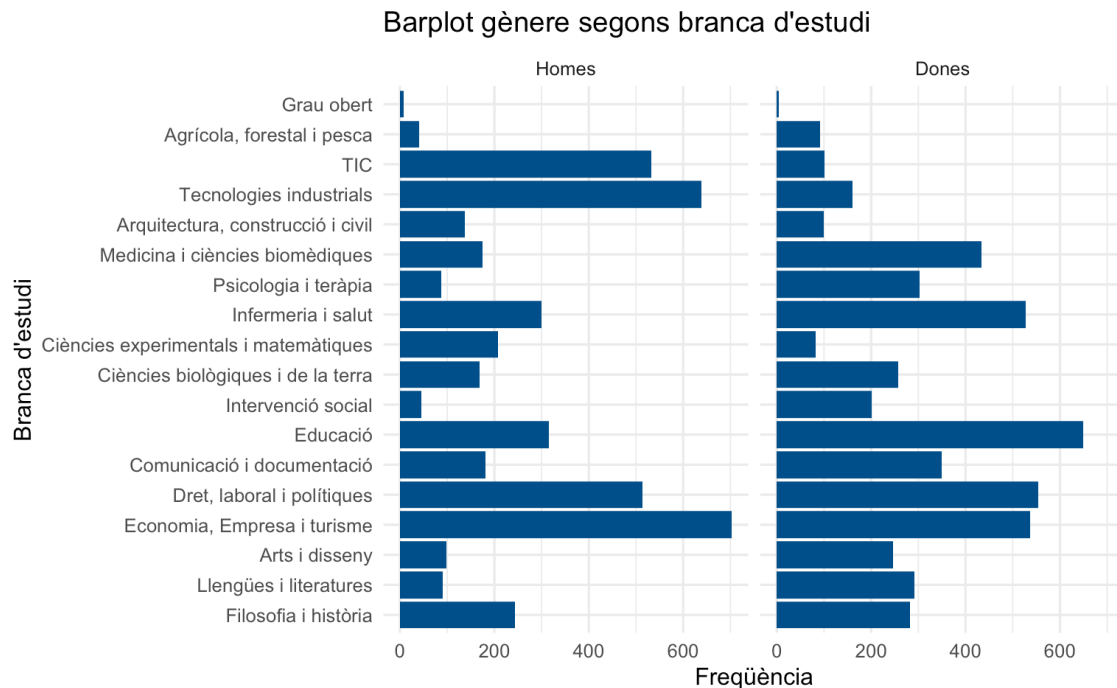


Figura 14: Gràfic de barres de les branques d'estudi en funció del gènere

La informació que aporta aquest gràfic és molt interessant ja que es poden distingir grans diferències entre els dos grups. El primer grup on existeix una diferència notable és en les branques de TIC i tecnologies industrials. Tendeixen a ser uns estudis molt poc freqüentats per dones i aquest gràfic ens ajuda a confirmar-ho. El segon grup on hi ha evidència de diferenciació és en les diverses branques de salut. En aquest grup s'engloba medicina, psicologia i teràpia i infermeria. És un sector conegut per la gran abundància de dones que hi treballen. Finalment, l'última branca que destaca és la d'educació on també hi participen una gran quantitat de dones.

Per a provar si aquests dos factors tenen relació es tornarà a fer el test d'independència de la  $\chi^2$  de Pearson. Seguim treballant amb un nivell de significació del 95%. Les dues hipòtesis són:

$H_0$ : Les variables són independents, per tant, la categoria de la variable "branca" no varia entre els diferents nivells de la variable "sexe".

$H_1$ : Les variables són dependents, per tant, la categoria d'una variable pot influir entre els diferents nivells de l'altra variable.

L'estadístic es construeix de la següent manera:

$$\chi^2 = \sum_{ij} \frac{(\text{observats}_{ij} - \text{esperats}_{ij})^2}{\text{esperats}_{ij}}$$

S'obtenen els següents resultats:

*Taula 5: Resultats del test d'independència*

Estadístic	Graus de llibertat	P-valor
1.398,4	17	2,2e-16

S'observa que el *p-valor* és molt més inferior a 0,05. Per tant, tenim suficient evidència per rebutjar la hipòtesi nul·la. És per això que podem dir que l'elecció del grau universitari estarà relacionat amb gènere de la persona.

#### 4. Relació entre branca d'estudi i nacionalitat

En aquest apartat procedirem a analitzar gràficament la relació entre la variable branca d'estudi i nacionalitat. Aquesta vegada és diferent a les anteriors, ja que ens trobem davant d'una situació que no afavoreix a la comparació de resultats. A l'observar la quantitat de gent amb nacionalitat espanyola i la restant es va veure que hi havia una diferència notable. De 9.655 estudiants, 8.969 eren amb nacionalitat espanyola (92,89%) i 686 amb nacionalitat no espanyola (7,11%). Per tant, s'ha trobat oportú treballar amb proporcions en comptes de treballar amb freqüències.

A continuació hi ha representat un gràfic de barres comparant les proporcions d'estudiants de les diferents branques diferenciats per la seva nacionalitat.

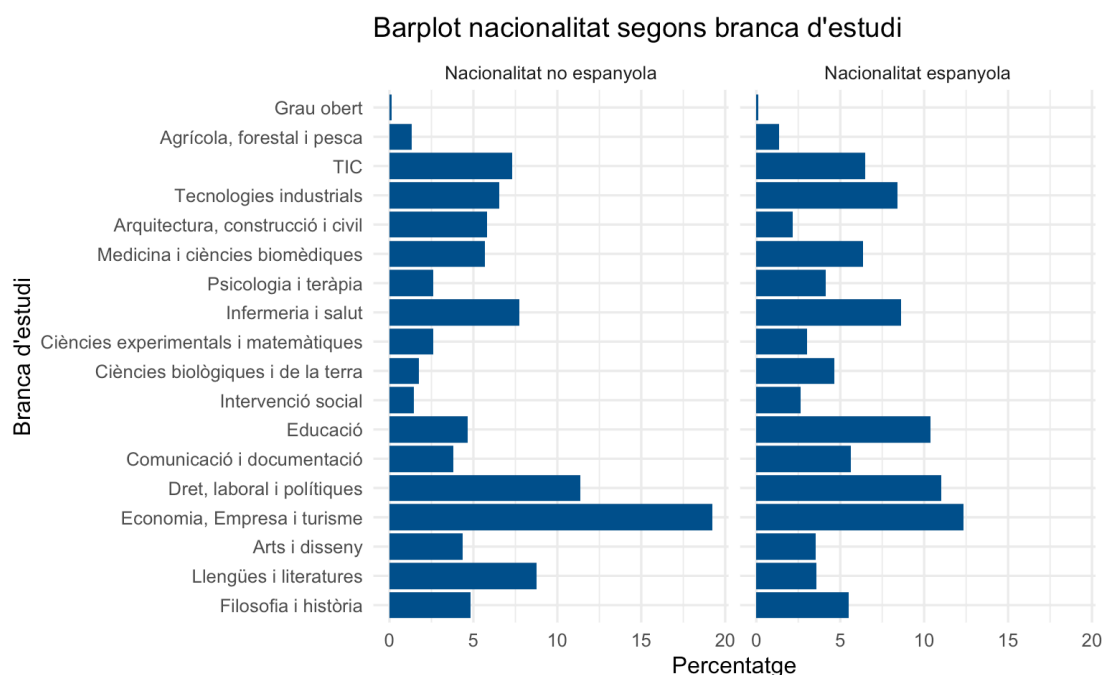


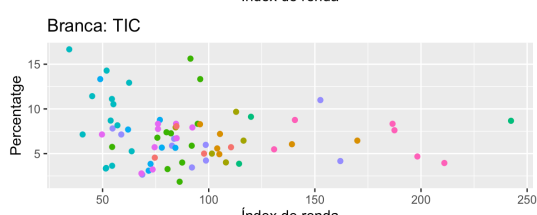
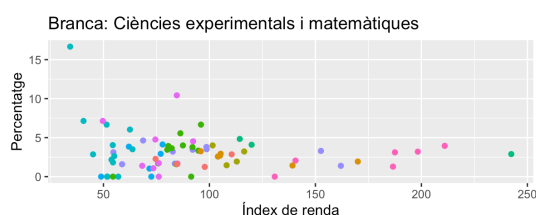
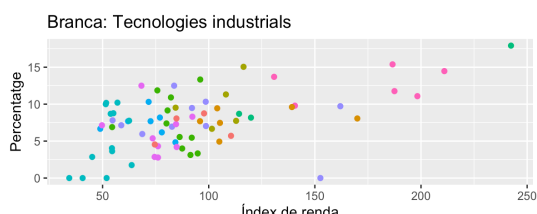
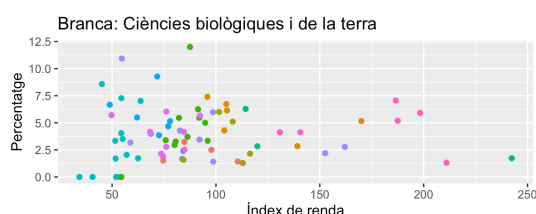
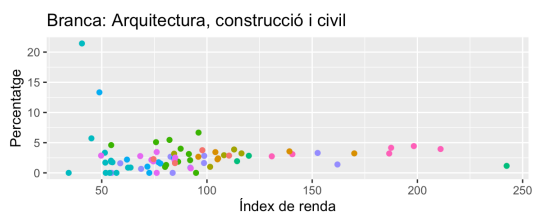
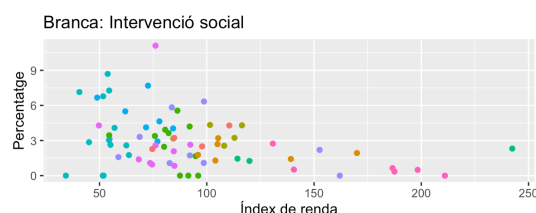
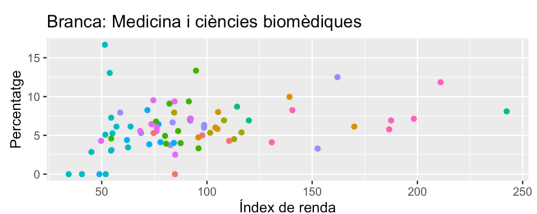
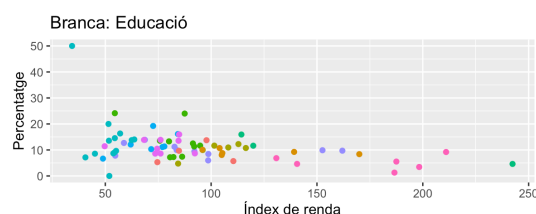
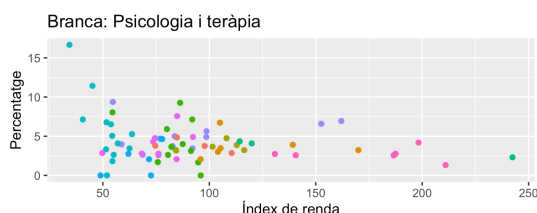
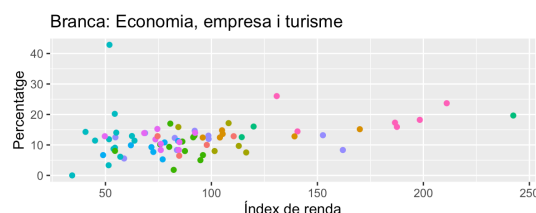
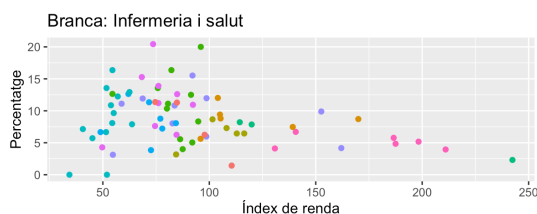
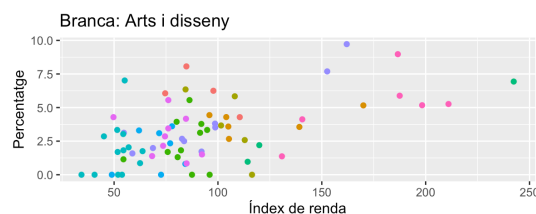
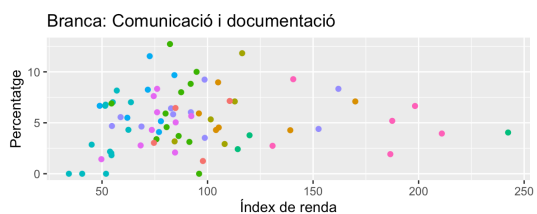
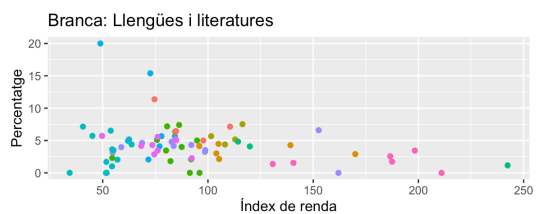
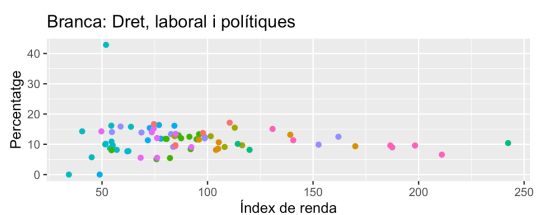
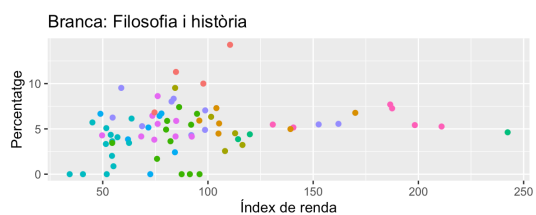
Figura 15: Gràfic de barres de les branques d'estudi en funció de la nacionalitat

En aquest gràfic es veu que no existeixen tantes diferències com en el cas de la variable anterior. Es pot destacar que els de nacionalitat no espanyola tenen un percentatge més elevat en les branques d'arquitectura, construcció i civil, llengües i literatura i sobretot en economia, empresa i turisme. Per altra banda, els estudiants de nacionalitat espanyola tenen un percentatge més alt en educació.

## 5. Relació entre branca d'estudi i Índex de renda familiar mitjana

Aquest apartat és una continuació del primer, però en aquest cas és més concret. S'analitzarà branca per branca, la relació entre la proporció d'estudiants per cada barri i l'índex de renda. L'anàlisi constarà de dues parts, primer es graficarà la informació i seguidament s'aplicarà el test de correlació de *Spearman* per conèixer el grau de correlació que tenen.

Tot i això, s'inicia amb un conjunt de gràfics de cada branca que mostra de manera visual la possible correlació que hi ha entre cada barri i la renda.



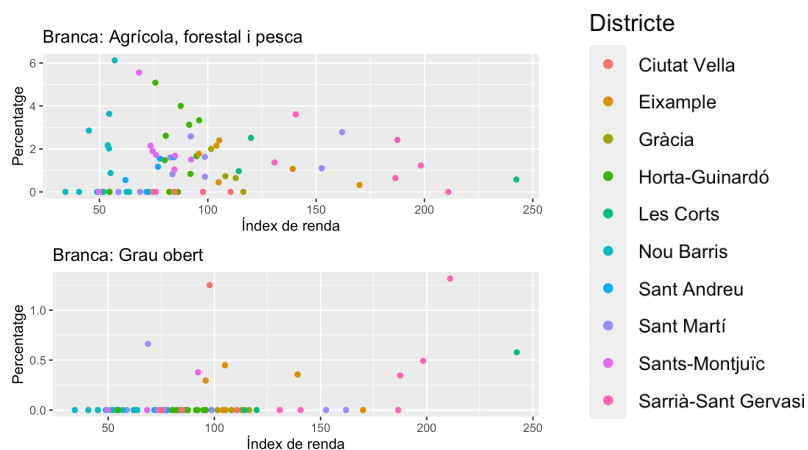


Figura 16: Conjunt de gràfics de correlació segons branques

Observant els gràfics de correlació es poden identificar quines són les branques que estan més correlacionades. Tot i així, a continuació es realitzarà el test de correlació de *Spearman* per a comprovar de forma estadística la relació entre les variables. Les hipòtesis que es contrastaran són:

$H_0: \rho = 0$  (no existeix relació)

$H_1: \rho \neq 0$  (sí existeix relació)

A continuació s'adjunta una taula amb un recull dels resultats per branques. S'hi representa la rho (coeficient de correlació de *Spearman*), el seu corresponent *p-valor* una columna amb un asterisc per ressaltar els tests que són significatius i rebutgen la hipòtesi nul·la.

Taula 6: Resultats test de correlació de *Spearman*

Branca d'estudi	Rho	P-valor	
Filosofia i història	0,271	0,019	*
Llengües i literatures	-0,131	0,266	
Arts i disseny	0,455	4,557e-05	*
Economia, empresa i turisme	0,345	0,003	*
Dret, laboral i polítiques	-0,032	0,788	
Comunicació i documentació	0,1326	0,260	
Educació	-0,320	0,005	*

<b>Intervenció social</b>	-0,342	0,003	*
<b>Ciències biològiques i de la terra</b>	0,081	0,493	
<b>Ciències experimentals i matemàtiques</b>	0,002	0,986	
<b>Infermeria i salut</b>	-0,240	0,040	*
<b>Psicologia i teràpia</b>	-0,119	0,312	
<b>Medicina i ciències biomèdiques</b>	0,340	0,003	*
<b>Arquitectura, construcció i civil</b>	0,270	0,020	*
<b>Tecnologies industrials</b>	0,405	0,0004	*
<b>TIC</b>	-0,087	0,46	
<b>Agrícola, forestal i pesca</b>	0,208	0,076	
<b>Grau obert</b>	0,375	0,0009	*

Per a fer-ho més visual, es representarà els diferents coeficients separats en dos grups: els que són significativament diferents de zero i els que no. El que es pretén amb aquest gràfic és ressaltar el nombre de coeficients que pertanyen en cada grup.

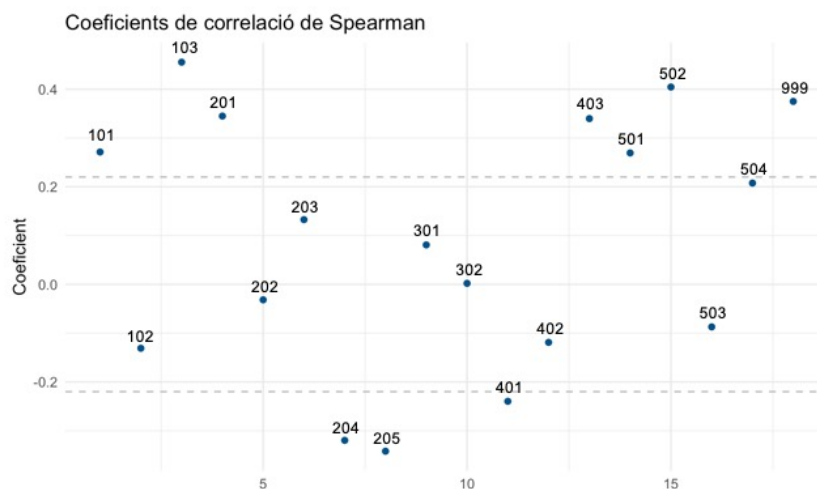


Figura 17: Gràfic de la distribució dels coeficients de Spearman

Finalment, obtenim un total de 10 branques sobre 18 amb una correlació significativament diferent de zero.



Pel que fa a les branques on s'ha obtingut un resultat significatiu:

En el cas de filosofia i història, art i disseny, economia, empresa i turisme, medicina i ciències biomèdiques, arquitectura, construcció i civil, tecnologies industrials i grau obert s'obté una relació positiva. Això s'interpreta que com més alt sigui l'índex de renda del barri, més percentatge d'estudiants estudiaran aquella branca. Es pot destacar en concret que tendeixen a ser branques del sector privat .

Per altra banda, apareixen les branques d'educació, intervenció social i infermeria i salut on el coeficient és negatiu. Això vol dir que com més baix sigui l'índex de renda, més persones cursen aquestes branques. Es pot observar que són branques on l'objectiu de la feina que es realitza és ajudar a terceres persones.

Pel que fa a les branques on s'ha obtingut un resultat no significatiu:

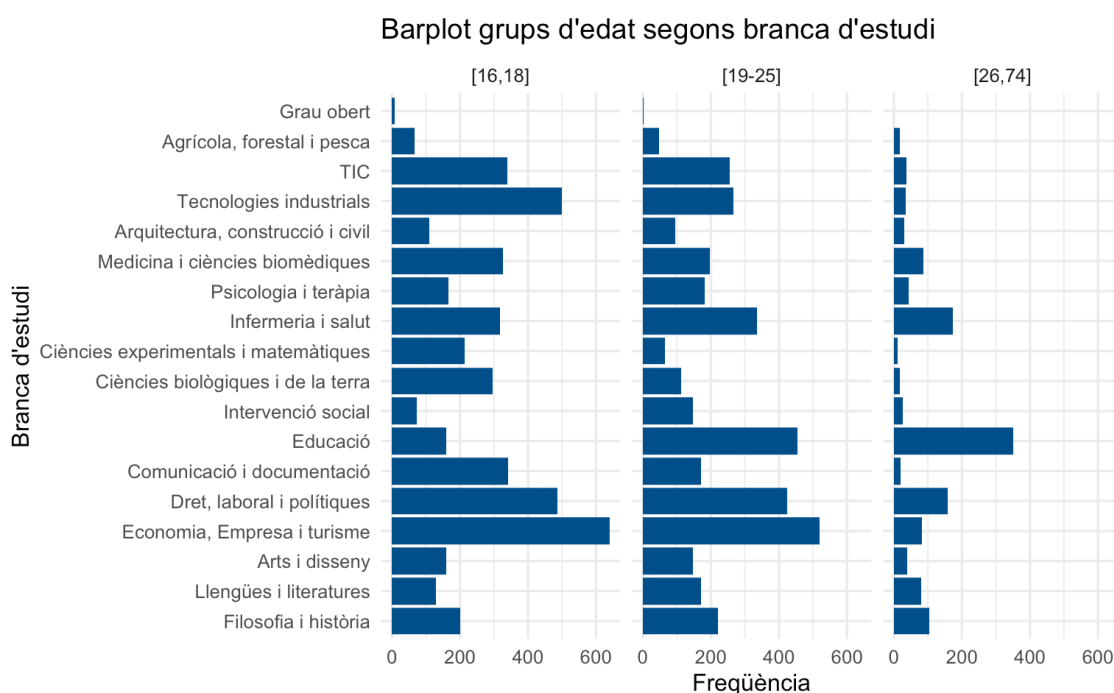
D'aquest grup de branques, es podria destacar en concret un grup que pot sorprendre els resultats. La branca de dret, laboral i polítiques, la qual és reconeguda com un sector amb alts ingressos, per tant es podria associar a llocs amb un índex de renda alt. En aquest cas aquesta suposició queda rebutjada.

## 6. Relació entre branca d'estudi i grups d'edat

En aquest últim apartat s'estudia la influència de l'edat a l'hora d'escollir les branques d'estudis universitaris. A continuació s'adjuntarà un gràfic amb les freqüències de persones segons la branca que escullen dividit entre tres franges d'edat.

Aquestes franges d'edat corresponen a:

- [16,18]: Alumnes que, si han seguit el transcurs natural dels estudis sense interrupcions, els pertoca l'inici a estudis universitaris. També s'inclou estudiants que han estat avançats en algun curs.
- [19,25]: Joves que accedeixen a un nou estudi universitari.
- [25,74]: Persones que accedeixen a un nou estudi universitari majors de 25 anys.



*Figura 18: Gràfic de barres de les branques d'estudi en funció de grups d'edat*

En aquest gràfic apareixen resultats força interessants. Es podria començar destacant que la freqüència de joves entre 16 i 18 anys que accedeixen a un grau universitari és similar a la de joves de 19 a 25. Per altre costat s'observa que la quantitat de persones majors de 15 anys es redueix molt.

En relació a les branques més populars entre els grups d'edat, destaquen tres grups que són els més freqüentats en els tres grups: educació, infermeria i salut i dret, laboral i polítiques.

Per altra banda, es pot observar que els més joves també tenen tendència a branques com TIC i tecnologies industrials (tendència a la branca científica), mentre que en el grup de persones de més edat tendeixen més a filosofia i història i llengües i literatures (tendència a la branca d'humanitats).

Finalment, s'ha de destacar la branca d'economia, empresa i turisme. És coneguda com una branca amb altes freqüències de persones, i ens ho confirmen els grups de 16 a 18 i de 19 a 25, mentre que en els majors de 25 no és de les que capta més interès.

## VIII. CONCLUSIONS

El propòsit principal d'aquest estudi era poder identificar quines eren les variables que podien influir més a l'hora de separar els futurs estudiants per les branques d'estudi proposades. La variable que havia captat més interès era la del barri on habiten. Tot i això, també era interessant poder conèixer i descriure el perfil dels individus de l'estudi, tant en variables de caràcter més personal, l'edat i el gènere, com en variables de caire econòmic, com ara l'índex de renda familiar mitjana.

Per poder arribar a estudiar els objectius marcats ha estat necessari un llarg tractament de la base de dades proporcionada. Això és degut a que les dades inicials no incorporaven quasi cap de les variables d'interès principal, per exemple el barri. El fet de poder assignar un barri residencial a cada estudiant ha suposat una gran dificultat, ja que només es tenia coneixença del seu codi postal. Per tant, s'ha fet un estudi exhaustiu de la distribució de la població dels barris i districtes de Barcelona, i de les diverses possibilitats de repartir els individus entre ells de la manera més correcta possible. Una vegada distribuïts els estudiants en els seus respectius barris, va ser necessari buscar informació per acabar de complementar la base de dades per poder realitzar de manera satisfactòria l'anàlisi amb les variables que es tenia en ment.

Tanmateix, quan es disposava d'una base de dades completa es va procedir a fer un anàlisi descriptiu univariant, seguit d'un conjunt de representacions de les variables sobre un mapa de la ciutat de Barcelona. Es va fer amb la intenció de fer més visual tota la informació que seria analitzada a posteriorment. A més, durant l'anàlisi bivariant es va detectar algun patró de distribució segons certes variables. Aquest patró va ser l'inici per a després comprovar, sobre aquestes mateixes variables, la influència que podien generar sobre l'individu a l'hora d'escollir una branca d'estudi en concret. En aquest treball es va voler centrar en l'anàlisi bivariant, ja que no es disposava de més informació per poder comprovar les relacions en més dimensions.

Amb els resultats obtinguts durant l'anàlisi es poden destacar varies coses interessants: per començar, la hipòtesi plantejada inicialment sobre si el barri on habita l'estudiant influeix en l'elecció de la branca d'estudis, queda demostrada que sí. A més, hi ha un seguit de branques que destaquen per la seva gran freqüència d'estudiants que hi volen accedir, com ara economia, empresa i turisme, educació i dret, laboral i polítiques. És un resultat que es podia preveure ja que en estudis anteriors s'ha demostrat que són els principals graus d'interès. Però en tot cas, aquestes freqüències són diferents segons el barri.

Dins d'aquesta relació també s'ha estudiat si l'índex de renda és una covariable important. Quan s'ha fet el test de manera general (percentatge d'estudiant per barri envers l'índex de renda), s'ha observat que sí i amb una correlació positiva, és a dir, que com més alt és l'índex de renda, més percentatge d'estudiants existeix per barri. Aquest

resultat no és inesperat, ja que per poder accedir a estudis universitaris és necessari un alt ingrés en l'economia familiar. Però quan s'ha entrat a distingir entre les diverses branques i barris, ha quedat demostrat que existeix una correlació per a algunes d'aquestes, mentre que per altres no. Aquests resultats són inesperats, ja que no s'havia plantejat la qüestió que en certes branques no pogués existir aquesta relació. En relació amb aquest resultat potser caldria investigació addicional per clarificar-ho. Com s'ha destacat en l'anàlisi, les variables amb una correlació positiva són aquelles que tendeixen a pertànyer al sector privat, mentre que les que han obtingut una correlació negativa, són les que tendeixen a ser de caire més públic i on l'objectiu de la feina és ajudar a terceres persones.

Pel que fa a les variables socials i demogràfiques, s'ha observat uns resultats esperats. Entrant en qüestions de gènere, s'ha comprovat que aquest factor influeix en la distribució dels estudiants segons les branques. Una altra variable que ha mostrat certes diferències pel que fa a freqüències d'estudiants per branques és l'edat. Es pot afirmar que depenent de l'edat, l'interès per cursar un tipus d'estudi o un altre és diferent. Per altra banda, en relació a la nacionalitat, les persones que provenen de fora d'Espanya mostren un alt interès en un nombre de branques en concret, que són: economia, empresa i turisme, dret, laboral i polítiques i llengües i literatures.

Recollint el més important, la hipòtesi principal plantejada ha estat acceptada i demostrada. També s'han detectat altres covariables i factors amb un cert grau d'influència. En aquest procés d'elecció, però, és molt important també l'entorn personal que envolta a cada estudiant. Per tant, cal destacar que per a cada persona, el grau d'influència de cada variable podria ser més o menys important depenent de la situació de cadascú.

## IX. BIBLIOGRAFIA I WEB GRAFIA

Agència per a la Qualitat del Sistema Universitari de Catalunya. Actualització 2020 del catàleg de titulacions d'AQU Catalunya [en línia]. Barcelona: AQU Catalunya, 09/02/2020. [Consulta: 06/08/2020]. Disponible a: <<http://www.aqu.cat/aqu/actualitat/noticies/53799401.html#.XkFzhy0-8dU>>

Agència per a la Qualitat del Sistema Universitari de Catalunya. Catàleg de titulacions [en línia]. Barcelona: AQU Catalunya, any 2020. [Consulta: 07/08/2020]. Disponible a: <<http://www.aqu.cat/estudis/catalegtitols.html#.XupMnC0zlZ1>>

Ajuntament de Barcelona. Distribució territorial de la renda familiar a Barcelona. 2000-2017. [en línia]. Barcelona: Ajuntament de Barcelona. Oficina Municipal de Dades, 01/01/2017. [Consulta: 17/07/2020]. Disponible a: <<https://www.bcn.cat/estadistica/catala/dades/economia/renda/rdfamiliar/evo/rfbaris.htm>>

Ajuntament de Barcelona, Gabinet Tècnic de Programació. Barcelona Economia. Distribució territorial de la renda familiar a Barcelona. 2016. [en línia]. Barcelona: Departament d'Estadística. Ajuntament de Barcelona, 01/01/2016. [Consulta: 17/07/2020]. Disponible a: <[https://ajuntament.barcelona.cat/barcelonaeconomia/sites/default/files/RFD\\_2016\\_BCN.pdf](https://ajuntament.barcelona.cat/barcelonaeconomia/sites/default/files/RFD_2016_BCN.pdf)>

Alonso, Diego. Cómo crear una composición de mapa con QGIS [en línia]. Valladolid: MappingGIS, any 2015. [Consulta: 14/12/2020]. Disponible a: <<https://mappinggis.com/2015/06/como-crear-una-composicion-de-mapa-con-qgis/>>

Canal Universitats. Informes i estadístiques de la preinscripció universitària [en línia]. Barcelona: Generalitat de Catalunya, any 2020. [Consulta: 14/05/2020]. Disponible a: <[http://universitats.gencat.cat/ca/altres\\_pagines/informe\\_i\\_estadistiques/informes\\_i\\_estad\\_pre/](http://universitats.gencat.cat/ca/altres_pagines/informe_i_estadistiques/informes_i_estad_pre/)>

Garret Grolemond. Data Visualization with ggplot2: Cheat Sheet [en línia]. Rstudio, 01/11/2019. [Consulta: 21/10/2020]. Disponible a: <<https://ggplot2.tidyverse.org>>

Institut d'Estadística de Catalunya. Renda familiar disponible bruta. Índex. Comarques i Aran, i àmbits. Metodologia [en línia]. Barcelona: IDESCAT, 30/04/2020. [Consulta: 10/05/2020]. Disponible a: <<https://www.idescat.cat/pub/?id=aec&n=941&m=m>>

Institut Municipal d'Informàtica. Unitats administratives de la ciutat de Barcelona [en línia]. Barcelona: Ajuntament de Barcelona, 06/07/2017. [Consulta: 4/12/2020]. Disponible a: <https://opendata-ajuntament.barcelona.cat/data/ca/dataset/20170706-districtes-barris> >

Instituto Nacional de Estadística. Relación de municipios y sus códigos por provincias. Últimos datos. [en línia]. INE, 07/02/2020. [Consulta: 2/05/2020]. Disponible a: <https://www.ine.es/daco/daco42/codmun/codmunmapa.htm> >

Oficina d'Accés a la Universitat. Notes de tall 1a assignació Juny 2019 [en línia]. Barcelona: Generalitat de Catalunya, 11/07/19. [Consulta: 08/08/2020]. Disponible a: [http://universitats.gencat.cat/web/.content/01\\_acces\\_i\\_admissio/preinscripcio/documentacio/Notes\\_i\\_places/Notes\\_tall\\_1a\\_assignacio\\_11-07-19\\_JUNY\\_2019.pdf](http://universitats.gencat.cat/web/.content/01_acces_i_admissio/preinscripcio/documentacio/Notes_i_places/Notes_tall_1a_assignacio_11-07-19_JUNY_2019.pdf)>

## X. ANNEX

### 1. Taules complementàries del registre

#### 1.1 Valors taula OCB

Opció batxillerat-COU.

Codi	Descripció
<b>A</b>	Científicotecnològica
<b>B</b>	Biosanitària
<b>C</b>	Ciències socials
<b>D</b>	Humanísticolingüística
<b>AB</b>	Opció doble Científicotecnològica-Biosanitària
<b>CD</b>	Opció doble Ciències socials-Humanísticolingüística
<b>1</b>	Científicotècnica
<b>2</b>	Ciències de la salut
<b>3</b>	Humanitats
<b>4</b>	Ciències socials
<b>5</b>	Arts
<b>12</b>	Opció doble Científicotecnològica-Ciències de la salut
<b>34</b>	Opció doble Humanitats-Ciències socials

*Taula 7: Codificació valors batxillerat-COU*

#### 1.2 Valors taula O25

Opció examen proves per als més grans de 25 anys.

Codi	Descripció
<b>1</b>	Científicotècnica
<b>2</b>	Ciències de la salut
<b>3</b>	Humanitats
<b>4</b>	Ciències socials
<b>5</b>	Arts
<b>12</b>	Opció doble Científicotècnica - Ciències de la salut (només en el cas de recuperació de dades acadèmiques anys anteriors opció CI (opció doble Científicotècnica - Ciències de la salut)
<b>34</b>	Opció doble Humanitats - Ciències socials
<b>35</b>	Opció doble Humanitats - Arts
<b>45</b>	Opció doble Ciències socials - Arts
<b>345</b>	Opció triple Humanitats - Ciències socials - Arts
<b>A</b>	Arts i humanitats
<b>B</b>	Ciències
<b>C</b>	Ciències de la salut
<b>D</b>	Ciències socials i jurídiques
<b>E</b>	Enginyeria i arquitectura
<b>BC</b>	Ciències - Ciències de la salut

<b>BCE</b>	Ciències - Ciències de la salut - Enginyeria i arquitectura
<b>BCD</b>	Ciències - Ciències de la salut - Ciències socials i jurídiques
<b>AD</b>	Arts i humanitats - Ciències socials i jurídiques
<b>CD</b>	Ciències de la salut - Ciències socials i jurídiques
<b>DE</b>	Ciències socials i jurídiques - Enginyeria i arquitectura

*Taula 8: Codificació valors d'opcions proves per a més de 25 anys*

### 1.3 Valors taula TM i CE

S'adjunta l'accés al Dropbox amb un arxiu Excel que conté la informació d'ambdues taules ja que són de dimensions molt grans.

[https://www.dropbox.com/s/tpzqx1bcmhls4eo/Taules\\_registre.xlsx?dl=0](https://www.dropbox.com/s/tpzqx1bcmhls4eo/Taules_registre.xlsx?dl=0)

### 1.4 Valors taula ES

Codificació d'estudis dels pares.

<b>Codi</b>	<b>Descripció</b>
<b>1</b>	Sense estudis
<b>2</b>	Estudis primaris
<b>3</b>	EGB o FP 1er grau
<b>4</b>	Batxillerat o FP 2n grau
<b>5</b>	Diplomat o Enginyer Tèc.
<b>6</b>	Dr., Llicenciat, Enginyer o Arquitecte
<b>7</b>	Altres / NS / NC

*Taula 9: Codificació valors del grau d'estudis dels pares*

### 1.5 Valors taula OC

Codificació d'ocupació dels pares.

<b>Codi</b>	<b>Descripció</b>
<b>1</b>	Director o Gerent d'empreses o Institucions Públiques
<b>2</b>	Tècnic o professió associada a tit. univ.
<b>3</b>	Treballador qualificat sect. agr. ramaderia o pesca
<b>4</b>	Treballador qualificat sect. industrial
<b>5</b>	Treballador qualificat sect. construcció i mineria
<b>6</b>	Treballador qualificat sect. serveis
<b>7</b>	Treballador no qualificat
<b>8</b>	Forces armades
<b>9</b>	Altres / NS / NC
<b>10</b>	Persones que no han tingut un treball remunerat

*Taula 10: Codificació valors d'ocupació dels pares*



## 1.6 Valors taula OR

### Codificació orientació

Codi	Descripció
1	Professors de secundària
2	Visita professor universitari al centre secundària
3	Salons d'Ensenyament
4	Persones que han fet aquests estudis
5	Altres / NS / NC
6	Amistats
7	Entorn familiar
8	Web "Què i per què estudiar a les universitats catalanes?" / Guia de l'ensenyament superior
9	Sortides professionals
10	Iniciativa pròpia/vocació
11	Interès en el contingut dels estudis

*Taula 11: Codificació valors d'orientació*

## 1.7 Valors taula CS

### Codificació dels criteris de selecció.

Codi	Descripció
1	Nota de tall
2	Durada de l'estudi
3	Facilitat o dificultat de l'estudi
4	Proximitat al domicili familiar
5	Expectatives professionals
6	Altres / NS / NC
7	Interès en el contingut dels estudis

*Taula 12: Codificació dels criteris de selecció*

## 1.8 Valors taula TI

### Codificació de la titulació dels estudiants.

Codi	Descripció
1	Llicenciat
2	Diplomat
3	Sense titulació amb 180 crèdits o més
4	Sense titulació amb menys de 180 crèdits
5	Altres / NS / NC
6	Doctor

*Taula 13: Codificació de la titulació dels estudiants*

## 2. Taula de codificació dels barris

Barri	Número del barri
el Camp de l'Arpa del Clot	1
el Clot	2
el Parc de la Llacuna del Poblenou	3
la Vila Olímpica del Poblenou	4
el Poblenou	5
Sant Martí de Provençals	6
Provençals del Poblenou	7
la Verneda i la Pau	8
el Besòs i el Maresme	9
Diagonal Mar i el Front Marítim del Poblenou	10
el Congrés i els Indians	11
Navas	12
la Sagrera	13
Sant Andreu	14
el Bon Pastor	15
Baró de Viver	16
la Trinitat Vella	17
Canyelles	18
Ciutat Meridiana	19
Torre Baró	20
Vallbona	21
la Trinitat Nova	22
Roquetes	23
la Guineueta	24
Verdun	25
la Prosperitat	26
Can Peguera	27
el Turó de la Peira	28
Porta	29
Vilapicina i la Torre Llobeta	30
Sant Genís dels Agudells	31
Montbau	32
la Vall d'Hebron	33
Horta	34
la Clota	35
la Teixonera	36
el Carmel	37
la Font d'en Fargues	38
Can Baró	39

<b>el Baix Guinardó</b>	<b>40</b>
<b>el Guinardó</b>	<b>41</b>
<b>Vallcarca i els Penitents</b>	<b>42</b>
<b>el Coll</b>	<b>43</b>
<b>la Salut</b>	<b>44</b>
<b>la Vila de Gràcia</b>	<b>45</b>
<b>el Camp d'en Grassot i Gràcia Nova</b>	<b>46</b>
<b>Sarrià</b>	<b>47</b>
<b>Sant Gervasi Bonanova</b>	<b>48</b>
<b>Sant Gervasi Galvany</b>	<b>49</b>
<b>Tres Torres</b>	<b>50</b>
<b>Putxet Farró</b>	<b>51</b>
<b>Vallvidrera Tibidabo les Planes</b>	<b>52</b>
<b>Pedralbes</b>	<b>53</b>
<b>Sant Ramon Maternitat</b>	<b>54</b>
<b>les Corts</b>	<b>55</b>
<b>la Nova Esquerra de l'Eixample</b>	<b>56</b>
<b>l'Antiga Esquerra de l'Eixample</b>	<b>57</b>
<b>la Dreta de l'Eixample</b>	<b>58</b>
<b>la Sagrada Família</b>	<b>59</b>
<b>el Fort Pienc</b>	<b>60</b>
<b>Sant Antoni</b>	<b>61</b>
<b>Sants-Badal</b>	<b>62</b>
<b>Sants</b>	<b>63</b>
<b>la Bordeta</b>	<b>64</b>
<b>Hostafrancs</b>	<b>65</b>
<b>la Font de la Guatlla</b>	<b>66</b>
<b>la Marina del Port</b>	<b>67</b>
<b>la Marina del Prat Vermell</b>	<b>68</b>
<b>el Poble-Sec</b>	<b>69</b>
<b>Parc de Monjuïc</b>	<b>70</b>
<b>Zona Franca-Port</b>	<b>71</b>
<b>el Raval</b>	<b>72</b>
<b>el Barri Gòtic</b>	<b>73</b>
<b>la Barceloneta</b>	<b>74</b>
<b>Sant Pere, Santa Caterina i la Ribera</b>	<b>75</b>

*Taula 14: Codificació dels barris*

### 3. Taula de codificació dels districtes

Districtes	Número del districte
Ciutat Vella	1
Eixample	2
Sants-Montjuïc	3
Les Corts	4
Sarrià-Sant Gervasi	5
Gràcia	6
Horta-Guinardó	7
Nou Barris	8
Sant Andreu	9
Sant Martí	10

Taula 15: Codificació dels districtes

### 4. Taula equivalència dels graus

Codi	Nom subàrea 1r nivell
101	Filosofia i Història
102	Llengües i Literatures
103	Arts i disseny
201	Economia, Empresa i Turisme
202	Dret, laboral i polítiques
203	Comunicació i Documentació
204	Educació
205	Intervenció Social
301	Ciències biològiques i de la terra
302	Ciències experimentals i matemàtiques
401	Infermeria i salut
402	Psicologia i teràpia
403	Medicina i Ciències Biomèdiques
501	Arquitectura, construcció i civil
502	Tecnologies industrials
503	TIC
504	Agrícola, forestal i pesca
999	Grau obert

Taula 16: Codificació codis AQU 1r nivell

## 5. Taula índex de renda familiar mitjana per barris

Barri	Índex de renda
el Camp de l'Arpa del Clot	82,7
el Clot	83,7
el Parc de la Llacuna del Poblenou	92,1
la Vila Olímpica del Poblenou	162
el Poblenou	98,6
Sant Martí de Provençals	68,7
Provençals del Poblenou	98,7
la Verneda i la Pau	58,8
el Besòs i el Maresme	54,7
Diagonal Mar i el Front Marítim del Poblenou	152,6
el Congrés i els Indians	71,7
Navas	84,3
la Sagrera	77,9
Sant Andreu	77
el Bon Pastor	62
Baró de Viver	72,6
la Trinitat Vella	48,9
Canyelles	54,5
Ciutat Meridiana	34,3
Torre Baró	45,1
Vallbona	51,9
la Trinitat Nova	40,6
Roquetes	57
la Guineueta	54,4
Verdun	53,8
la Prosperitat	55,1
Can Peguera	51,5
el Turó de la Peira	51,7
Porta	62,5
Vilapicina i la Torre Llobeta	63,7
Sant Genís dels Agudells	86,3
Montbau	82,2
la Vall d'Hebron	87,5
Horta	80,6
la Clota	91,4
la Teixonera	75,8
el Carmel	54,5
la Font d'en Fargues	94,8
Can Baró	96

el Baix Guinardó	92
el Guinardó	80,1
Vallcarca i els Penitents	112,9
el Coll	84,4
la Salut	116,4
la Vila de Gràcia	101,5
el Camp d'en Grassot i Gràcia Nova	108,1
Sarrià	186,6
Sant Gervasi Bonanova	187,5
Sant Gervasi Galvany	198,3
Tres Torres	211
Putxet Farró	140,6
Vallvidrera Tibidabo les Planes	130,8
Pedralbes	242,4
Sant Ramon Maternitat	114,3
les Corts	119,9
la Nova Esquerra de l'Eixample	105,3
l'Antiga Esquerra de l'Eixample	139,2
la Dreta de l'Eixample	170
la Sagrada Família	95,9
el Fort Pienc	105
Sant Antoni	104
Sants-Badal	74,5
Sants	92,3
la Bordeta	73,6
Hostafrancs	84,6
la Font de la Guatlla	84,9
la Marina del Port	68,3
la Marina del Prat Vermell	49,7
el Poble-Sec	76,1
Parc de Monjuïc	76,1
Zona Franca-Port	49,7
el Raval	74,6
el Barri Gòtic	110,5
la Barceloneta	84,8
Sant Pere, Santa Caterina i la Ribera	97,8

Taula 17: Índex de renda familiar mitjana per barri

## 6. Recodificació de les variables

Tret d'estudi	Tipus	Descripció
<b>ID_ALUMNE</b>	Categòrica	Identificador per cada alumne.
<b>NACIONALITAT</b>	Categòrica	Nacionalitat de cada estudiant. <ul style="list-style-type: none"> <li>• "1": Espanya</li> <li>• "2": Resta d'Europa</li> <li>• "3": Àsia</li> <li>• "4": Amèrica del Nord</li> <li>• "5": Amèrica del Sud</li> <li>• "6": Oceania</li> <li>• "7": Àfrica</li> </ul>
<b>CP</b>	Categòrica	Codi postal del domicili familiar o residència habitual.
<b>SEXE</b>	Categòrica	<ul style="list-style-type: none"> <li>• "0": Home</li> <li>• "1": Dona</li> </ul>
<b>edatT</b>	Numèrica	Edat teòrica.
<b>VIA ACCES</b>	Categòrica	<ul style="list-style-type: none"> <li>• "0": PAU o assimilats</li> <li>• "2": Llicenciats, diplomats o assimilats</li> <li>• "4": CFGS o assimilats</li> <li>• "7": PAU amb carrera començada</li> <li>• "8": CFGS o assimilats amb carrera començada</li> <li>• "9": MG25 anys</li> <li>• "10": MG40 anys amb experiència laboral</li> </ul>
<b>OPC COU/BATXILLERAT/MG25</b>	Categòrica	Opcions Batxillerat/COU. Segons valors de la taula OCB. Opcions més grans de 25 anys. Segons valors de la taula O25.
<b>ANY QUAL</b>	Categòrica	Any de qualificació de les PAU. <ul style="list-style-type: none"> <li>• "1": Entre 1967 i 2003</li> <li>• "2": Entre 2004 i 2009</li> <li>• "3": Entre 2010 i 2011</li> <li>• "4": L'any 2012</li> <li>• "5": L'any 2013</li> <li>• "6": L'any 2014</li> <li>• "7": L'any 2015</li> <li>• "8": L'any 2018</li> </ul>
<b>NOTA ACCES PAU/EXPEDIENT CFGS</b>	Numèrica	Nota d'accés a les PAU / Nota expedient CFGS.
<b>NOTA_ADMISSIO</b>	Numèrica	Nota d'admissió a la universitat.

<b>SIST_EDU_ES</b>	Categòrica	País amb sistema educatiu estranger sense PAU. <ul style="list-style-type: none"> <li>• “0”: No</li> <li>• “1”: Sí</li> </ul>
<b>PREFE1</b>	Categòrica	Codi de cada grau segons el llistat vigent per cada convocatòria.
<b>ULT_CARR_COM</b>	Categòrica	Última carrera començada. Segons valors de la taula TM.
<b>ULT_CURS_FIN</b>	Categòrica	Any de la última carrera finalitzada (“AAAA”). <ul style="list-style-type: none"> <li>• “1”: Entre 1969 i 2009</li> <li>• “2”: Entre 2010 i 2013</li> <li>• “3”: Entre 2014 i 2015</li> <li>• “4”: L’any 2016</li> </ul>
<b>CENTRE ASSIG</b>	Categòrica	Centre assignat. Segons valor de la taula CE.
<b>ORDRE ASSIG</b>	Numèrica	Ordre d’assignació del centre d’estudis.
<b>CONCOR</b>	Categòrica	Concordança entre preferència 1 i centre assignat. <ul style="list-style-type: none"> <li>• “0”: No</li> <li>• “1”: Sí</li> </ul>
<b>ESTU_PARE</b>	Categòrica	Estudis acabats del pare. Segons valors de la taula ES.
<b>ESTU_MARE</b>	Categòrica	Estudis acabats de la mare. Segons valors de la taula ES.
<b>OCUP_PARE</b>	Categòrica	Ocupació del pare. Segons valors de la taula OC.
<b>OCUP_MARE</b>	Categòrica	Ocupació de la mare. Segons valors de la taula OC.
<b>TREB_REM</b>	Categòrica	Treball remunerat de l’alumne. <ul style="list-style-type: none"> <li>• “1”: Treballa menys de 15 hores setmanals</li> <li>• “2”: Treballa 15 o més hores setmanals</li> <li>• “3”: No treballa</li> </ul>
<b>ORIENTACIO</b>	Categòrica	Orientació a escollir els estudis universitaris. Segons valors de la taula OR.
<b>ELEC_EST</b>	Categòrica	Aspectes que s’han tingut en compte en l’elecció d’estudi. Segons valors de la taula CS.
<b>TIP_CENTRE</b>	Categòrica	Tipus de centre on s’ha estudiat l’últim any de secundària. <ul style="list-style-type: none"> <li>• “1”: Públic</li> <li>• “2”: Privat</li> <li>• “3”: Concertat</li> </ul>



<b>TIP_CEN_FP</b>	Categòrica	Tipus de centre on s'ha estudiat l'últim ant de CFGS o FP2. <ul style="list-style-type: none"> <li>• "1": Públic</li> <li>• "2": Privat</li> <li>• "3": Concertat</li> </ul>
<b>EST_ANT</b>	Categòrica	S'ha cursat estudis universitaris anteriors. <ul style="list-style-type: none"> <li>• "0": No</li> <li>• "1": Sí</li> </ul>
<b>TITULACIO</b>	Categòrica	Titulació o nivell assolit. Segons valor de la taula TI.
<b>CENTRE MATRIC</b>	Categòrica	Codi del centre matriculat. Segons valors de la taula CE.
<b>BARRI</b>	Categòrica	Barri assignat a cada alumne.
<b>renda16</b>	Numèrica	Renda mitjana familiar per barri de l'any 2016.
<b>PREFE1_1</b>	Categòrica	Primera preferència. Graus universitaris segons codificació AQU nivell 1.
<b>PREFE1_2</b>	Categòrica	Primera preferència. Graus universitaris segons codificació AQU nivell 2.
<b>CENTRE.ASSIG_1</b>	Categòrica	Centre assignat. Graus universitaris segons codificació AQU nivell 1.
<b>CENTRE.ASSIG_2</b>	Categòrica	Centre assignat. Graus universitaris segons codificació AQU nivell 2.
<b>CENTRE.MATRIC_1</b>	Categòrica	Centre matriculat. Graus universitaris segons codificació AQU nivell 1.
<b>CENTRE.MATRIC_2</b>	Categòrica	Centre matriculat. Graus universitaris segons codificació AQU nivell 2.

*Taula 18: Recodificació de les variables*

## 7. Graus que no apareixien en la classificació

- 41051: Estudis globals – grup 101 (Filosofia i història)
- 41050: Programa de grau Obert – grup 999 (Grau Obert)
- 10120: Investigació privada – grup 202 (Dret, laboral i polítiques)
- 41053: Multidisciplinari en tecnologia i negocis "Tecnocampus" – grup 201 (Economia, empresa i turisme)

- 21080: Nanociència i nanotecnologia – grup 301 (Ciències biològiques i de la terra)

- 11076: Seguretat – grup 202 (Dret, laboral i polítiques)

## 8. Descriptiva de les variables escollides

### 8.1 Variable Nacionalitat

Districte	Espanya	Resta d'Europa	Àsia	Amèrica del Nord	Amèrica del Sud	Oceania	Àfrica
<b>Ciutat Vella</b>	82,80%	6,40%	6,10%	0,90%	3,20%	0,00%	0,60%
<b>Eixample</b>	93,20%	2,20%	1,80%	0,20%	2,20%	0,10%	0,30%
<b>Gràcia</b>	93,80%	2,80%	1,10%	0,20%	1,80%	0,00%	0,20%
<b>Horta-Guinardó</b>	93,90%	1,70%	1,10%	0,80%	2,10%	0,00%	0,40%
<b>Les Corts</b>	94,60%	2,10%	1,40%	0,60%	1,00%	0,00%	0,30%
<b>Nou Barris</b>	91,40%	2,00%	1,30%	1,10%	3,80%	0,00%	0,40%
<b>Sant Andreu</b>	93,40%	1,40%	2,00%	0,10%	2,60%	0,00%	0,50%
<b>Sant Martí</b>	91,50%	3,00%	1,80%	0,30%	2,80%	0,00%	0,60%
<b>Sants-Montjuïc</b>	92,20%	2,60%	2,80%	0,20%	2,20%	0,00%	0,10%
<b>Sarrià-Sant Gervasi</b>	95,40%	2,70%	0,80%	0,40%	0,70%	0,00%	0,10%

Taula 19: Percentatges per districtes variable nacionalitat

### 8.2 Variable Sexe

Districte	Dona	Home
<b>Ciutat Vella</b>	57,30%	42,70%
<b>Eixample</b>	52,40%	47,60%
<b>Gràcia</b>	53,80%	46,20%
<b>Horta-Guinardó</b>	54,40%	45,60%
<b>Les Corts</b>	51,70%	48,30%
<b>Nou Barris</b>	57,30%	42,70%
<b>Sant Andreu</b>	55,40%	44,60%
<b>Sant Martí</b>	55,00%	45,00%
<b>Sants-Montjuïc</b>	52,90%	47,10%
<b>Sarrià-Sant Gervasi</b>	50,00%	50,00%

Taula 20: Percentatges per districtes variable sexe

## 9. Taula amb percentatge d'estudiants per barri

Barri	Percentatge
el Camp de l'Arpa del Clot	1,937
el Clot	1,243
el Parc de la Llacuna del Poblenou	1,201
la Vila Olímpica del Poblenou	0,746
el Poblenou	1,906
Sant Martí de Provençals	1,564
Provençals del Poblenou	1,471
la Verneda i la Pau	1,305
el Besòs i el Maresme	0,663
Diagonal Mar i el Front Marítim del Poblenou	0,943
el Congrés i els Indians	1,005
Navas	1,284
la Sagrera	2,009
Sant Andreu	1,771
el Bon Pastor	1,885
Baró de Viver	0,269
la Trinitat Vella	0,155
Canyelles	0,57
Ciutat Meridiana	0,062
Torre Baró	0,363
Vallbona	0,073
la Trinitat Nova	0,145
Roquetes	0,508
la Guineueta	1,025
Verdun	0,476
la Prosperitat	1,181
Can Peguera	0,311
el Turó de la Peira	0,611
Porta	1,201
Vilapicina i la Torre Llobeta	1,181
Sant Genís dels Agudells	0,559
Montbau	0,57
la Vall d'Hebron	0,259
Horta	1,585
la Clota	0,331
la Teixonera	0,611
el Carmel	0,901
la Font d'en Fargues	0,621
Can Baró	0,311

<b>el Baix Guinardó</b>	2,465
<b>el Guinardó</b>	2,103
<b>Vallcarca i els Penitents</b>	1,605
<b>el Coll</b>	0,653
<b>la Salut</b>	0,963
<b>la Vila de Gràcia</b>	3,107
<b>el Camp d'en Grassot i Gràcia Nova</b>	2,838
<b>Sarrià</b>	1,616
<b>Sant Gervasi Bonanova</b>	2,993
<b>Sant Gervasi Galvany</b>	4,205
<b>Tres Torres</b>	0,787
<b>Putxet Farró</b>	2,009
<b>Vallvidrera Tibidabo les Planes</b>	0,756
<b>Pedralbes</b>	1,792
<b>Sant Ramon Maternitat</b>	2,144
<b>les Corts</b>	3,294
<b>la Nova Esquerra de l'Eixample</b>	3,884
<b>l'Antiga Esquerra de l'Eixample</b>	2,91
<b>la Dreta de l'Eixample</b>	3,211
<b>la Sagrada Família</b>	3,501
<b>el Fort Pienc</b>	2,31
<b>Sant Antoni</b>	2,413
<b>Sants-Badal</b>	1,088
<b>Sants</b>	2,745
<b>la Bordeta</b>	0,963
<b>Hostafrancs</b>	0,994
<b>la Font de la Guatlla</b>	1,233
<b>la Marina del Port</b>	0,746
<b>la Marina del Prat Vermell</b>	0,725
<b>el Poble-Sec</b>	1,201
<b>Parc de Monjuïc</b>	0,373
<b>Zona Franca-Port</b>	0
<b>el Raval</b>	1,367
<b>el Barri Gòtic</b>	0,725
<b>la Barceloneta</b>	0,642
<b>Sant Pere, Santa Caterina i la Ribera</b>	0,829

*Taula 21: Percentatge d'estudiants per barri*

## 10. Codi utilitzat en R

### 10.1 SCRIPT: Depuració del codi postal

#### Funció:

```
```{r}
cortar <- function(x,y){ # x: vector de valors i NA ; y: número
  v <- c()
  for(i in 1:y){
    v <- c(v, x[i])
  }
  return(v)
}

posiciones <- function(x){
  v <- c(x[1])
  for(i in 2:length(x)){
    aux <- v[i-1] + x[i]
    v <- c(v, aux)
  }
  return(v)
}
```
```

#### Dades:

```
```{r}
assignats <- readxl:: read_xlsx("SolicitantsBCN.xlsx")
CP <- readxl:: read_xlsx("CP.xlsx", sheet = 1)
Barris <- readxl:: read_xlsx("CP.xlsx", sheet = 2)
```
```

**Procés d'assignar a cada alumne amb el seu CP corresponent, el barri que pertoca segons la divisió.**

```
```{r}
set.seed(123)
```

```

# Lloc de recoll·lecta de dades
alumnebarri <- data.frame(assignats$`4 ID_ALUMNE`)

v1 <- CP$CP
N <- length(v1)
M <- nrow(assignats)
for(i in 1:N){
  v2 <- c()
  aux <- v1[i]

  # vector amb les posicions dels alumnes amb el mateix cp
  for(j in 1:M){
    if(aux == assignats$`14 CP`[j]){
      v2 <- c(v2, j)
    }
  }

  # tractament per cada cp (aux)
  total <- length(v2) # total alumnes per barri
  totalgrups <- c(CP[i,])

  # nombre de barris que pertany el CP
  nop <- (sum(is.na(totalgrups) == F) - 2)/2

  # vector amb porcions x barri
  nporcions <- c(totalgrups$Porcio1, totalgrups$Porcio2, totalgrups$Porcio3,
totalgrups$Porcio4, totalgrups$Porcio5, totalgrups$Porcio6, totalgrups$Porcio7)

  # vector dels barris que pertany
  nbarris <- c(totalgrups$Barri1, totalgrups$Barri2, totalgrups$Barri3, totalgrups$Barri4,
totalgrups$Barri5, totalgrups$Barri6, totalgrups$Barri7)

  # creació de valors per a calcular les particions dels alumnes de cada CP
  npor <- cortar(nporcions, nop)

```

```

divisions <- sum(npor)
random <- data.frame()
alumnes_aleatori <- sample(1:total, total, replace = F)
particions <- c((total/divisions)*npor)
particions_exacte <- trunc(particions)

# creació de vctor de 1,...,7 per assignar grups
aux2 <- c()
for(k in 1:length(particions_exacte)){
  tamany <- particions_exacte[k]
  aux3 <- rep(k, tamany)
  aux2 <- c(aux2, aux3)
}

# tractament alumnes de divisió no exacta + assignació aleatoria a un barri dels 1,...,7
if(sum(particions_exacte) != total){
  noexact <- total - sum(particions_exacte)
  alea <- c(sample(1:nop, noexact, replace = F))
  aux2 <- c(aux2, alea)
}

# Bdd on hi ha posició alumnes aleatoris i barri assignat
df <- data.frame(alumnes_aleatori, aux2)
dford <- df[with(df, order(df$alumnes_aleatori)), ]
dftotal <- cbind(dford, v2)

# assignació barri a la bdd final
if(total != 0){
  for(l in 1:nrow(dftotal)){
    for(n in 1:nop){
      if(dftotal$aux2[l] == n){
        vv <- dftotal$v2[l]
        alumnebarri[vv,2] <- nbarris[n]
      }
    }
  }
}

```

```

    }
  }
}
}

```

```

colnames(alumnebarri) <- c("Alumne", "Barri")
write.csv2(alumnebarri, "AlumnesBarris.csv", row.names = F)
...

```

## 10.2 SCRIPT: Tractament de la base de dades

### ## Tractament general

#### Dades:

```

```{r}
library(readxl)
soli <- read_excel("~/Desktop/TFG/Dades/SolicitantsBCN_Net.xlsx")
barris <- read_excel("~/Desktop/TFG/Dades/AlumnesBarrisAct.xlsx")
...

```

#### Assignació de barri a cada alumne corresponent:

```

```{r}
BARRI <- c()
for(i in 1:nrow(soli)){
  for(j in 1:nrow(barris)){
    if(soli$ID_ALUMNE[i] == barris$Alumne[j]){
      BARRI[i] <- barris$Barri[j]
    }
  }
}
soli <- data.frame(soli, BARRI)
save.image("tractament_barri.RData")
...

```

#### Assignació variable barri - renda mitjana:



```

```{r}
load("tractament_barri.RData")

renda <- read_excel("~/Desktop/TFG/Dades/Renda - barris.xlsx")

renda16 <- c()
for(i in 1:nrow(soli)){
  for(j in 1:nrow(renda)){
    if(soli$BARRI[i] == renda$Num_Barri[j]){
      renda16 <- c(renda16, renda$renda16[j])
    }
  }
}
soli <- data.frame(soli, renda16)
```

```

#### **Assignació codis AQU(2) - PREFE1**

```

```{r}
dduni <- readxl::read_xlsx("~/Desktop/TFG/Dades/BDD Codis Universitats.xlsx", sheet =
1)

# Modificar tipus de variables de la bdd dduni
str(dduni)
dduni$Codi <- as.character(dduni$Codi)
dduni$Codi_AQU1 <- as.character(dduni$Codi_AQU1)

# Comprovació si hi ha algun codi de la bdd soli que no coincideix amb algun de la bdd
dduni

comp <- c()
v <- c()

for(i in soli$PREFE1){
  for(j in dduni$Codi){
    if(i == j){
      comp <- c(comp, TRUE)
    }
  }
}

```

```

    } else {
      comp <- c(comp, FALSE)
    }
  }
  if(sum(comp == TRUE) == 1){
    v <- c(v, T)
  } else {v <- c(v, F)}
  comp <- c()
}

s <- c(which(v == F))
codis <- c(soli$PREFE1[s]) # comprovació: hauria de donar 0

# Assignació
PREFE1_1Nivell <- c()
PREFE1_2Nivell <- c()
for(i in 1:nrow(soli)){
  for(j in 1:nrow(dduni)){
    if(soli$PREFE1[i] == dduni$Codi[j]){
      PREFE1_1Nivell <- c(PREFE1_1Nivell, dduni$Codi_AQU1[j])
      PREFE1_2Nivell <- c(PREFE1_2Nivell, dduni$Codi_AQU2[j])
    }
  }
}

```

```

soli <- data.frame(soli, PREFE1_1Nivell, PREFE1_2Nivell)

```

```

soli$PREFE1_1Nivell <- as.character(soli$PREFE1_1Nivell)
soli$PREFE1_2Nivell <- as.character(soli$PREFE1_2Nivell)
```

```

### **Assignació codis AQU(2) - CENTRE.ASSIG**

```

```{r}

```

```

# Comprovació si hi ha algun codi de la bdd soli que no coincideix amb algun de la bdd
dduni

```

```
comp2 <- c()
```

```
v2 <- c()
```

```
for(i in soli$CENTRE.ASSIG){
```

```
  for(j in dduni$Codi){
```

```
    if(i == j){
```

```
      comp2 <- c(comp2, TRUE)
```

```
    } else {
```

```
      comp2 <- c(comp2, FALSE)
```

```
    }
```

```
  }
```

```
  if(sum(comp2 == TRUE) == 1){
```

```
    v2 <- c(v2, T)
```

```
  } else {v2 <- c(v2, F)}
```

```
  comp2 <- c()
```

```
}
```

```
s2 <- c(which(v2 == F))
```

```
codis2 <- c(soli$CENTRE.ASSIG[s2]) # comprovació: hauria de donar 0
```

```
# Assignació
```

```
CENTRE.ASSIG_1 <- c()
```

```
CENTRE.ASSIG_2 <- c()
```

```
for(i in 1:nrow(soli)){
```

```
  for(j in 1:nrow(dduni)){
```

```
    if(soli$CENTRE.ASSIG[i] == dduni$Codi[j]){
```

```
      CENTRE.ASSIG_1 <- c(CENTRE.ASSIG_1, dduni$Codi_AQU1[j])
```

```
      CENTRE.ASSIG_2 <- c(CENTRE.ASSIG_2, dduni$Codi_AQU2[j])
```

```
    }
```

```
  }
```

```
  if(soli$CENTRE.ASSIG[i] == 0){
```

```
    CENTRE.ASSIG_1 <- c(CENTRE.ASSIG_1, 0)
```

```
    CENTRE.ASSIG_2 <- c(CENTRE.ASSIG_2, 0)
```

```
}  
}
```

```
soli <- data.frame(soli, CENTRE.ASSIG_1, CENTRE.ASSIG_2)
```

```
soli$CENTRE.ASSIG_1 <- as.character(soli$CENTRE.ASSIG_1)
```

```
soli$CENTRE.ASSIG_2 <- as.character(soli$CENTRE.ASSIG_2)
```

```
...
```

### **Assignació codis AQU(2) - CENTRE.MATRIC**

```
```{r}
```

```
# Comprovació si hi ha algun codi de la bdd soli que no coincideix amb algun de la bdd  
dduni
```

```
comp3 <- c()
```

```
v3 <- c()
```

```
for(i in soli$CENTRE.MATRIC){
```

```
  for(j in dduni$Codi){
```

```
    if(i == j){
```

```
      comp3 <- c(comp3, TRUE)
```

```
    } else {
```

```
      comp3 <- c(comp3, FALSE)
```

```
    }
```

```
  }
```

```
  if(sum(comp3 == TRUE) == 1){
```

```
    v3 <- c(v3, T)
```

```
  } else {v3 <- c(v3, F)}
```

```
  comp3 <- c()
```

```
}
```

```
s3 <- c(which(v3 == F))
```

```
codis3 <- c(soli$CENTRE.MATRIC[s3]) # comprovació: hauria de donar 0
```

```
# Assignació
```

```
CENTRE.MATRIC_1 <- c()
```

```

CENTRE.MATRIC_2 <- c()
for(i in 1:nrow(soli)){
  for(j in 1:nrow(dduni)){
    if(soli$CENTRE.MATRIC[i] == dduni$Codi[j]){
      CENTRE.MATRIC_1 <- c(CENTRE.MATRIC_1, dduni$Codi_AQU1[j])
      CENTRE.MATRIC_2 <- c(CENTRE.MATRIC_2, dduni$Codi_AQU2[j])
    }
  }
  if(soli$CENTRE.MATRIC[i] == "00000"){
    CENTRE.MATRIC_1 <- c(CENTRE.MATRIC_1, 0)
    CENTRE.MATRIC_2 <- c(CENTRE.MATRIC_2, 0)
  }
}

```

```

soli <- data.frame(soli, CENTRE.MATRIC_1, CENTRE.MATRIC_2)

```

```

save.image("dades_finals.RData")
...

```

### **Recodificació:**

```

```{r}

```

```

load("dades_finals.RData")

```

```

# VARIABLES CATEGÒRIQUES

```

```

# Recodificació variable SEXE

```

```

for(i in 1:nrow(soli)){
  if(soli$SEXE[i] == "H"){
    soli$SEXE[i] <- 0
  } else if(soli$SEXE[i] == "D"){
    soli$SEXE[i] <- 1
  }
}
}

```

```
# Recodificació variable SIST_EDU_ES
for(i in 1:nrow(soli)){
  if(is.na(soli$SIST_EDU_ES)[i] == "FALSE"){
    if(soli$SIST_EDU_ES[i] == "N"){
      soli$SIST_EDU_ES[i] <- 0
    } else if(soli$SIST_EDU_ES[i] == "S"){
      soli$SIST_EDU_ES[i] <- 1
    }
  }
}
}
```

```
# Recodificació variable CONCOR
for(i in 1:nrow(soli)){
  if(is.na(soli$CONCOR)[i] == "FALSE"){
    if(soli$CONCOR[i] == "N"){
      soli$CONCOR[i] <- 0
    } else if(soli$CONCOR[i] == "S"){
      soli$CONCOR[i] <- 1
    }
  }
}
}
```

```
# Recodificació variable NACIONALITAT
nacionalitat <- read_excel("~/Desktop/TFG/Dades/nacionalitat.xlsx")
for(i in 1:nrow(soli)){
  for(j in 1:nrow(nacionalitat)){
    if(soli$NACIONALITAT[i] == nacionalitat$Pais[j]){
      soli$NACIONALITAT[i] <- nacionalitat$Codific[j]
    }
  }
}
}
```

```
# Recodificació variable ANY.QUAL
```

```

for(i in 1:nrow(soli)){
  if (soli$ANY.QUAL[i] >= 1967 && soli$ANY.QUAL[i] <= 2003) { soli$ANY.QUAL[i] <- 1
  } else if (soli$ANY.QUAL[i] >= 2004 && soli$ANY.QUAL[i] <= 2009) { soli$ANY.QUAL[i]
<- 2
  } else if (soli$ANY.QUAL[i] == 2010 || soli$ANY.QUAL[i] == 2011) { soli$ANY.QUAL[i] <-
3
  } else if (soli$ANY.QUAL[i] == 2012) { soli$ANY.QUAL[i] <- 4
  } else if (soli$ANY.QUAL[i] == 2013) { soli$ANY.QUAL[i] <- 5
  } else if (soli$ANY.QUAL[i] == 2014) { soli$ANY.QUAL[i] <- 6
  } else if (soli$ANY.QUAL[i] == 2015) { soli$ANY.QUAL[i] <- 7
  } else if (soli$ANY.QUAL[i] == 2016) { soli$ANY.QUAL[i] <- 8 }
}

```

# Recodificació variable ULT\_CURS\_FIN

```

for(i in 1:nrow(soli)){
  if(is.na(soli$ULT_CURS_FIN[i]) == FALSE){
    if (soli$ULT_CURS_FIN[i] >= 1969 && soli$ULT_CURS_FIN[i] <= 2009) {
soli$ULT_CURS_FIN[i] <- 1
    } else if (soli$ULT_CURS_FIN[i] >= 2010 && soli$ULT_CURS_FIN[i] <= 2013) {
soli$ULT_CURS_FIN[i] <- 2
    } else if (soli$ULT_CURS_FIN[i] == 2014 || soli$ULT_CURS_FIN[i] == 2015) {
soli$ULT_CURS_FIN[i] <- 3
    } else if (soli$ULT_CURS_FIN[i] == 2016) { soli$ULT_CURS_FIN[i] <- 4 }
  }
}

```

# VARIABLES NUMÈRIQUES

# Recodificació variable DATA\_NAIX

```

edatT <- c()
soli$DATA_NAIX <- as.numeric(as.character(soli$DATA_NAIX))
for(i in 1:nrow(soli)){
  aux <- 2016 - soli$DATA_NAIX[i]

```

```

    edatT <- c(edatT, aux)
}

soli <- data.frame(soli[,1:4], edatT, soli[,5:37])

# Recodificació variable NOTA.ACCES.PAU.EXPEDIENT.CFGS

soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS <-
as.numeric(soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS)

#for(i in 1:nrow(soli)){
#      if (soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] >= 5000 &&
soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] < #6000) {
soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] <- 1
#    } else if (soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] >= 6000 &&
#soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] < 7000) {
soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] <- 2
#    } else if (soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] >= 7000 &&
#soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] < 8000) {
soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] <- 3
#    } else if (soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] >= 8000 &&
#soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] < 9000) {
soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] <- 43
#    } else if (soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] >= 9000 &&
#soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] <= 10000) {
soli$NOTA.ACCES.PAU.EXPEDIENT.CFGS[i] <- 5
#  }
#}

# Recodificació variable NOTA_ADMISSIO

soli$NOTA_ADMISSIO <- as.numeric(soli$NOTA_ADMISSIO)

# for(i in 1:nrow(soli)){
#   if (soli$NOTA_ADMISSIO[i] >= 5000 && soli$NOTA_ADMISSIO[i] < 6000) {
soli$NOTA_ADMISSIO[i] <- 1
#   } else if (soli$NOTA_ADMISSIO[i] >= 6000 && soli$NOTA_ADMISSIO[i] < 7000) {
soli$NOTA_ADMISSIO[i] <- 2

```



```

# } else if (soli$NOTA_ADMISSION[i] >= 7000 && soli$NOTA_ADMISSION[i] < 8000) {
soli$NOTA_ADMISSION[i] <- 3

# } else if (soli$NOTA_ADMISSION[i] >= 8000 && soli$NOTA_ADMISSION[i] < 9000) {
soli$NOTA_ADMISSION[i] <- 4

# } else if (soli$NOTA_ADMISSION[i] >= 9000 && soli$NOTA_ADMISSION[i] < 10000) {
soli$NOTA_ADMISSION[i] <- 5

# } else if (soli$NOTA_ADMISSION[i] >= 10000 && soli$NOTA_ADMISSION[i] < 11000) {
soli$NOTA_ADMISSION[i] <- 6

# } else if (soli$NOTA_ADMISSION[i] >= 11000 && soli$NOTA_ADMISSION[i] < 12000) {
soli$NOTA_ADMISSION[i] <- 7

# } else if (soli$NOTA_ADMISSION[i] >= 12000 && soli$NOTA_ADMISSION[i] < 13000) {
soli$NOTA_ADMISSION[i] <- 8

# } else if (soli$NOTA_ADMISSION[i] >= 13000 && soli$NOTA_ADMISSION[i] <= 14000) {
soli$NOTA_ADMISSION[i] <- 9

# }

#}

...

```

#### **Guardar la base de dades:**

```

```{r}

write.csv2(soli, file = "~/Desktop/TFG/Dades/Soli.csv", row.names = F)

...

```{r}

rm(list = ls())

...

```

#### **## Missings**

```

```{r}

dd <- read.csv2("~/Desktop/TFG/Dades/Soli.csv")

...

```

#### **Taula amb recompte de missings:**

```

```{r}

taula <- c()

for(i in 1:ncol(dd)){
  a <- sum(is.na(dd[,i]))
  taula <- c(taula, a)
}

```

```
missings <- data.frame("Variable" = colnames(dd), "Num Missings" = taula)
```

```
write.csv2(missings, file = "missings.csv", row.names = F)
```

```
...
```

### *10.3 SCRIPT: Descriptiva*

#### **# Lectura de la bdd**

```
```{r}
```

```
setwd("~/Desktop/TFG/Dades")
```

```
dd <- read.csv2("Soli.csv")
```

```
...
```

#### **# Descriptiva univariant**

```
```{r}
```

```
## DESCRIPTIVA UNIVARIANT DE LES VARIABLES CATEGÒRIQUES
```

```
library(knitr)
```

```
library(descr)
```

```
library(ggplot2)
```

```
library(RColorBrewer)
```

```
library(grid)
```

```
library(dplyr)
```

```
library(GGally)
```

```
library(Hmisc)
```

```
library(corrplot)
```

```
library(PerformanceAnalytics)
```

```
library(gridExtra)
```

```
library(kableExtra)
```

```
# Nacionalitat
```

```
kable(freq(dd$NACIONALITAT, plot = F))
```

```
data1 <- dd
```

```

data1$NACIONALITAT[data1$NACIONALITAT == 2 | data1$NACIONALITAT == 3 |
data1$NACIONALITAT == 4 | data1$NACIONALITAT == 5 | data1$NACIONALITAT == 6 |
data1$NACIONALITAT == 7] <- 0

# CP
kable(freq(dd$CP, plot = F))

# Sexe
kable(freq(dd$SEXE, plot = F))

# Data naixement
kable(freq(dd$edatT, plot = F))

# VIA.ACCES
kable(freq(dd$VIA.ACCES, plot = F))

# OPC.CO.U.BATXILLERAT.MG25
kable(freq(dd$OPC.CO.U.BATXILLERAT.MG25, plot = F))

# ANY.QUAL
kable(freq(dd$ANY.QUAL, plot = F))

# SIST_EDU_ES
kable(freq(dd$SIST_EDU_ES, plot = F))

# ULT_CARR_COM
kable(freq(dd$ULT_CARR_COM != "NA", plot = F)[,1:2])
t <- freq(dd$ULT_CARR_COM != "NA", plot = F)[,1:2]
t <- as.table(t[1:2,1])

#ULT_CURS_FIN
kable(freq(dd$ULT_CURS_FIN != "NA", plot = F)[,1:2])
q <- freq(dd$ULT_CURS_FIN != "NA", plot = F)[,1:2]
q <- as.table(q[1:2,1])

#CONCOR
kable(freq(dd$CONCOR, plot = F))

#ESTU_PARE
kable(freq(dd$ESTU_PARE, plot = F))

#ESTU_MARE
kable(freq(dd$ESTU_MARE, plot = F))

#OCUP_PARE
kable(freq(dd$OCUP_PARE, plot = F))

#OCUP_MARE

```

```

kable(freq(dd$OCUP_MARE, plot = F))
#TREB_REM
kable(freq(dd$TREB_REM, plot = F))
#ORIENTACIO
kable(freq(dd$ORIENTACIO, plot = F))
#ELEC_EST
kable(freq(dd$ELEC_EST, plot = F))
#TIP_CENTRE
kable(freq(dd$TIP_CENTRE, plot = F))
#TIP_CEN_FP
kable(freq(dd$TIP_CEN_FP, plot = F))
#TITULACIO
kable(freq(dd$TITULACIO, plot = F))
#BARRI
kable(freq(dd$BARRI, plot = F))

```

## ## VARIABLES CATEGÒRIQUES

```

p1 <- ggplot(as.data.frame(table(data1$NACIONALITAT)), aes(x= c("Resta del món",
"Espanya"), y=Freq), labels) + geom_bar(stat = "identity", fill = "#00538C") + labs(title
="Barplot Nacionalitat", x = "Nacionalitat", y = "Freqüència") + theme_minimal()

```

```

p2 <- ggplot(as.data.frame(table(dd$SEXE)), aes(x= Var1, y=Freq), labels) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title ="Barplot Sexe", x = "Sexe", y =
"Freqüència") + theme_minimal()

```

```

p3 <- ggplot(as.data.frame(table(dd$VIA.ACCES)), aes(x= Var1, y=Freq), labels) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title ="Barplot Via d'accés", x = "Via
d'accés", y = "Freqüència") + theme_minimal()

```

```

p4 <- ggplot(as.data.frame(table(dd$ANY.QUAL)), aes(x= Var1, y=Freq), labels) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title ="Barplot Any de qualificació",
x = "Any qualificació", y = "Freqüència") + theme_minimal()

```

```

p5 <- ggplot(as.data.frame(t), aes(x= c("Sí", "No"), y=Freq)) + geom_bar(stat = "identity",
fill = "#00538C") + labs(title ="Barplot grau començat", x = "Carrera començada", y =
"Freqüència") + theme_minimal()

```

```
p6 <- ggplot(as.data.frame(q), aes(x= c("Sí", "No"), y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot grau acabat", x = "Carrera acabada", y
= "Freqüència") + theme_minimal()
```

```
p7 <- ggplot(as.data.frame(table(dd$CONCOR)), aes(x= Var1, y=Freq)) + geom_bar(stat
= "identity", fill = "#00538C") + labs(title = "Barplot Concordança", x = "Concordança", y
= "Freqüència") + theme_minimal()
```

```
p8 <- ggplot(as.data.frame(table(dd$ESTU_PARE)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Estudis pare", x =
"Nivell d'estudis", y = "Freqüència") + theme_minimal()
```

```
p9 <- ggplot(as.data.frame(table(dd$ESTU_MARE)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Estudis mare", x =
"Nivell d'estudis", y = "Freqüència") + theme_minimal()
```

```
p10 <- ggplot(as.data.frame(table(dd$OCUP_PARE)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Ocupació del pare", x
= "Treball", y = "Freqüència") + theme_minimal()
```

```
p11 <- ggplot(as.data.frame(table(dd$OCUP_MARE)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Ocupació de la mare",
x = "Treball", y = "Freqüència") + theme_minimal()
```

```
p12 <- ggplot(as.data.frame(table(dd$TREB_REM)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Treball", x = "Hores
de treball", y = "Freqüència") + theme_minimal()
```

```
p13 <- ggplot(as.data.frame(table(dd$ORIENTACIO)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Orientació", x =
"Mètode d'orientació", y = "Freqüència") + theme_minimal()
```

```
p14 <- ggplot(as.data.frame(table(dd$ELEC_EST)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Elecció de l'estudi", x
= "Aspectes tinguts en compte", y = "Freqüència") + theme_minimal()
```

```
p15 <- ggplot(as.data.frame(table(dd$TIP_CENTRE)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Centre d'estudi", x =
"Tipus de centre", y = "Freqüència") + theme_minimal()
```

```
p16 <- ggplot(as.data.frame(table(dd$TIP_CEN_FP)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Centre d'estudi FP", x
= "Tipus de centre FP", y = "Freqüència") + theme_minimal()
```

```
p17 <- ggplot(as.data.frame(table(dd$TITULACIO)), aes(x= Var1, y=Freq)) +
geom_bar(stat = "identity", fill = "#00538C") + labs(title = "Barplot Titulació", x = "Nivell
assolit", y = "Freqüència") + theme_minimal()
```

```
grid.arrange(p1, p2, p3, p4, p5, p6, p7, p8, p9, nrow=3, top = "DESCRIPTIVA UNIVARIANT
DE LES VARIABLES CATEGÒRIQUES")
```

```
grid.arrange(p10, p11, p12, p13, p14, p15, p16, p17, nrow=3)
```

```
## VARIABLES NUMÈRIQUES
```

```
# Edat teòrica
```

```
p18 <- qplot(dd$edatT,data=dd, geom = "density", main= paste("Density plot edat
teòrica"), xlab= "Edat") + theme_minimal()
```

```
# Nota.acces.pau.expedient.cfgs
```

```
p19 <- qplot(dd$NOTA.ACCES.PAU.EXPEDIENT.CFGS,data=dd, geom = "density", main=
paste("Density plot nota accés PAU"),xlab= "Nota") + theme_minimal()
```

```
# Nota d'admissió
```

```
p20 <- qplot(dd$NOTA_ADMISSIO,data=dd, geom = "density", main= paste("Density
plot nota d'admissió"), xlab = "Nota") + theme_minimal()
```

```
# Renda16
```

```
p22 <- qplot(dd$renda16,data=dd, geom = "density", main= paste("Density plot
renda16"), xlab= "Renda mitjana") + theme_minimal()
```

```
grid.arrange(p18, p19, p20, p22, nrow=2, top = "DESCRIPTIVA UNIVARIANT DENSITY
PLOT DE LES VARIABLES NUMÈRIQUES")
```

```
# Taula de dades
```

```
analisi <- function(x){
```

```
df <- data.frame("Min" = summary(x)[[1]], "Q1" = summary(x)[[2]], "Mediana" =
summary(x)[[3]], "Mitjana" = summary(x)[[4]], "Q3" = summary(x)[[5]], "Max" =
summary(x)[[6]])
}
```

```
df <- rbind(analisi(dd$edatT), analisi(dd$NOTA.ACCESSION.PAU.EXPEDIENT.CFGS),
analisi(dd$NOTA_ADMISSION), analisi(dd$ORDRE.ASSIG), analisi(dd$renda16))
rownames(df) <- c("Edat Teòrica", "Nota accés expedient", "Nota admissió", "Ordre
assignació", "Renda16")
kable(df,caption = "Taula descriptiva variables numèriques") %>%
kable_styling(c("striped", "bordered"))
```

```
#Boxplot
```

```
# Edat Teòrica
```

```
p23 <- ggplot(data = dd, aes(x = "", y = dd$edatT)) + geom_boxplot(fill = "#00538C") +
ylab("Edat") + theme_minimal() + labs(title = "Boxplot edat teòrica")
```

```
# Nota.acces.pau.expedient.cfgs
```

```
p24 <- ggplot(data = dd, aes(x = "", y = dd$NOTA.ACCESSION.PAU.EXPEDIENT.CFGS)) +
geom_boxplot(fill = "#00538C") + ylab("Nota") + theme_minimal() + labs(title = "Boxplot
nota accés pau expedient")
```

```
# Nota d'admissió
```

```
p25 <- ggplot(data = dd, aes(x = "", y = dd$NOTA_ADMISSION)) + geom_boxplot(fill =
"#00538C") + ylab("Nota") + theme_minimal() + labs(title = "Boxplot nota d'admissió")
```

```
# Renda16
```

```
p27 <- ggplot(data = dd, aes(x = "", y = dd$renda16)) + geom_boxplot(fill = "#00538C")
+ ylab("Renda mitjana") + theme_minimal() + labs(title = "Boxplot renda familiar
mitjana")
```

```
grid.arrange(p23, p24, p25, p27, nrow=2, top = "DESCRIPTIVA UNIVARIANT BOXPLOT DE
LES VARIABLES NUMÈRIQUES")
```

```
# Histograma
```

```
# Ordre assignació
```

```
ggplot(as.data.frame(table(dd$ORDRE.ASSIG)), aes(x= Var1, y=Freq)) +
geom_histogram(binwidth = 0, stat = "identity", fill = "#00538C") + labs(title = "Barplot
Ordre d'assignació", x = "Ordre", y = "Freqüència") + theme_minimal()
```

```
```
```

### # Elecció de variables i depuració de missings

```
```{r}
```

```
dd2 <- data.frame( "ID" = dd$ID_ALUMNE, "nac" = dd$NACIONALITAT, "sexe" =
dd$SEXE, "edatT" = dd$edatT, "nota" = dd$NOTA.ACCES.PAU.EXPEDIENT.CFGS,
"treball" = dd$TREB_REM, "barri" = dd$BARRI, "RFM" = dd$renda16, "P1N" =
dd$PREFE1_1Nivell, "P2N" = dd$PREFE1_2Nivell)
```

```
nas <- c(which(is.na(dd2$P1N) == T))
```

```
dd2 <- dd2[-nas,]
```

```
sum(is.na(dd2) == T) # comprovació
```

```
```
```

### # Posar variable districte a la base de dades

```
```{r}
```

```
barris <- readxl:: read_xlsx("~/Desktop/TFG/Dades/AlumnesBarrisAct.xlsx", sheet = 2)
```

```
districte <- c()
```

```
for(i in 1:nrow(dd2)){
```

```
  for(j in 1:nrow(barris)){
```

```
    if(dd2$barri[i] == barris$Num_Barri[j]){
```

```
      districte <- c(districte, barris$Districte[j])
```

```
    }
```

```
  }
```

```
}
```

```
dd2 <- data.frame(dd2, districte)
```

```
write.csv2(dd2, file = "~/Desktop/TFG/Dades/def.csv", row.names = F)
```

```
```
```

### # Variable resposta

```
```{r}
```

```
d1 <- subset(dd2, dd2$districte == 1)
```

```
d2 <- subset(dd2, dd2$districte == 2)
```

```
d3 <- subset(dd2, dd2$districte == 3)
```



```
d4 <- subset(dd2, dd2$districte == 4)
d5 <- subset(dd2, dd2$districte == 5)
d6 <- subset(dd2, dd2$districte == 6)
d7 <- subset(dd2, dd2$districte == 7)
d8 <- subset(dd2, dd2$districte == 8)
d9 <- subset(dd2, dd2$districte == 9)
d10 <- subset(dd2, dd2$districte == 10)
```

```
d1_1 <- ggplot(as.data.frame(table(d1$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Ciutat Vella", x = "Branca d'estudi", y =
"Freqüència") + theme_minimal() + coord_flip()
```

```
d1_2 <- ggplot(as.data.frame(table(d2$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Eixample", x = "Branca d'estudi", y =
"Freqüència") + theme_minimal() + coord_flip()
```

```
d1_3 <- ggplot(as.data.frame(table(d3$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Sants - Montjuïc", x = "Branca d'estudi",
y = "Freqüència") + theme_minimal() + coord_flip()
```

```
d1_4 <- ggplot(as.data.frame(table(d4$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Les Corts", x = "Branca d'estudi", y =
"Freqüència") + theme_minimal() + coord_flip()
```

```
d1_5 <- ggplot(as.data.frame(table(d5$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Sarrià - Sant Gervasi", x = "Branca
d'estudi", y = "Freqüència") + theme_minimal() + coord_flip()
```

```
d1_6 <- ggplot(as.data.frame(table(d6$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Gràcia", x = "Branca d'estudi", y =
"Freqüència") + theme_minimal() + coord_flip()
```

```
d1_7 <- ggplot(as.data.frame(table(d7$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Hosta - Guinadró", x = "Branca
d'estudi", y = "Freqüència") + theme_minimal() + coord_flip()
```

```
d1_8 <- ggplot(as.data.frame(table(d8$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Nou Barris", x = "Branca d'estudi", y =
"Freqüència") + theme_minimal() + coord_flip()
```

```
d1_9 <- ggplot(as.data.frame(table(d9$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat =
"identity", fill = "#00538C") + labs(title = "Barplot Sant Andreu", x = "Branca d'estudi", y
= "Freqüència") + theme_minimal() + coord_flip()
```

```
d1_10 <- ggplot(as.data.frame(table(d10$P1N)), aes(x= Var1, y=Freq)) + geom_bar(stat
= "identity", fill = "#00538C") + labs(title = "Barplot Sant Martí", x = "Branca d'estudi", y
= "Freqüència") + theme_minimal() + coord_flip()
```

```
grid.arrange(d1_1, d1_2, nrow=1, top = "DESCRIPTIVA BIVARIANT BARPLOT VARIABLE
RESPOSTA")
```

```
grid.arrange(d1_3, d1_4, nrow=1)
```

```
grid.arrange(d1_5, d1_6, nrow=1)
```

```
grid.arrange(d1_7, d1_8, nrow=1)
```

```
grid.arrange(d1_9, d1_10, nrow=1)
```

```
...
```

### # Variables explicatives

```
```{r, eval= F}
```

```
taula_edat <- data.frame(sort(unique(dd2$districte)))
```

```
taula_nota <- data.frame(sort(unique(dd2$districte)))
```

```
for(i in 1:10){
```

```
  aux <- subset(dd2, dd2$districte == i)
```

```
  taula_edat[i,2] <- mean(aux$edatT)
```

```
  taula_nota[i,2] <- mean(aux$nota)
```

```
  taula_edat[i,3] <- median(aux$edatT)
```

```
  taula_nota[i,3] <- median(aux$nota)
```

```
}
```

```
distr <- readxl::read_xlsx("~/Desktop/TFG/Dades/Renda - barris.xlsx", sheet = 2)
```

```
mitjanes <- data.frame(taula_edat[,1], taula_edat[,2], taula_nota[,2], distr$renda16)
```

```
write.csv2(mitjanes, "descr_pred.csv", row.names = F)
```

```
...
```

### Gràfic bivariant gènere - branques

```
```{r}
t <- as.data.frame(table(dd2$P1N, by = dd2$sexe))
levels(t$by) <- c("Homes","Dones")
levels(t$Var1) <- c("Filosofia i història", "Llengües i literatures", "Arts i disseny",
"Economia, Empresa i turisme", "Dret, laboral i polítiques", "Comunicació i
documentació", "Educació", "Intervenció social", "Ciències biològiques i de la terra",
"Ciències experimentals i matemàtiques", "Infermeria i salut", "Psicologia i teràpia",
"Medicina i ciències biomèdiques", "Arquitectura, construcció i civil", "Tecnologies
industrials", "TIC", "Agrícola, forestal i pesca", "Grau obert")
ggplot(t, aes(Var1, Freq)) + geom_bar(stat = "identity", fill = "#00538C") + labs(title
="Barplot gènere segons branca d'estudi", x = "Branca d'estudi", y = "Freqüència") +
theme_minimal() + coord_flip() + facet_grid(~by)
```
```

### 10.4 SCRIPT: Proporcions i anàlisi de dades

#### Dades:

```
```{r}
setwd("~/Desktop/TFG/Dades")
dd <- read.csv2("def.csv")
distr <- readxl::read_xlsx("AlumnesBarrisAct.xlsx", sheet = 3)
barris <- readxl::read_xlsx("AlumnesBarrisAct.xlsx", sheet = 2)
```
```

#### Posar total d'alumnes per districte

```
```{r}
for(i in 1:10){
  sub <- subset(dd, districte == i)
  distr[i,3] <- nrow(sub)
}
colnames(distr) <- c("districtes", "num", "Total")
```
```

#### Total d'alumnes per barri actualitzat

```
```{r}
for(i in 1:75){
  sub <- subset(dd, barri == i)
```

```

barris[i,5] <- nrow(sub)
}
colnames(barris) <- c("Barri", "Num_barri", "Total", "Districte", "Total_act")
...

```

### Funció

```

```{r}
analisi <- function(x,v,k,total,nom){ "x = vector amb llistat de branques d'alumnes d'un
sol barri/districte; v = vector amb llistat de les branques ordenades; aux = nom del
barri/districte; total = num d'alumnes"

taula <- data.frame(v)
for(i in 1:length(v)){
  suma <- 0
  for(j in 1:length(x)){
    if(v[i] == x[j]){
      suma <- suma + 1
    }
  }
  taula[i,2] <- suma
}
taula[,3] <- c(rep(k, 18))
taula[,4] <- c(rep(nom, 18))
taula[,5] <- taula[,2]/total
taula[,6] <- taula[,2]/total*100

colnames(taula) <- c("branca", "recompte", "districte", "nom_distr", "proporcio",
"percentatge")
return(taula)
}
...

```

### Taula proporcions de x alumnes que fan la branca Y del districte Z

```

```{r}
#bdd districtes
districtes <- data.frame()

aux <- as.numeric(c(1:10))
noms <- distr$districtes

```

```

v <- sort(unique(dd$P1N))
for(i in 1:10){ # per cada districte
  k <- aux[i]
  nom <- noms[i]
  sub <- data.frame(subset(dd, districte == k))
  x <- c(sub$P1N)
  total <- distr$Total[k]
  tau <- as.data.frame(analisi(x,v,k,total,nom))
  districtes <- rbind(districtes,tau)
}
...

```

### **Taula proporcions de x alumnes que fan la branca Y del barri Z**

```

```{r}
#bdd barris
barris_prop <- data.frame()

aux <- as.numeric(c(1:70,72:75)) # treiem el 71 perquè no té cap estudiant
noms <- barris$Barri[-71] # treiem zona franca
v <- sort(unique(dd$P1N))
for(i in 1:74){ # per cada barri menys zona franca
  k <- aux[i]
  nom <- noms[i]
  sub <- data.frame(subset(dd, barri == k))
  x <- c(sub$P1N)
  total <- barris$Total_act[k]
  tau <- as.data.frame(analisi(x,v,k,total,nom))
  barris_prop <- rbind(barris_prop,tau)
}
colnames(barris_prop) <- c("branca", "recompte", "barris", "nom_barri", "proporcio",
"percentatge")
...

```

### **Guardar excels**

```

```{r}
write.csv2(districtes, "Proporcions_districtes.csv", row.names = F)

```

```
write.csv2(barris_prop, "Proporcions_barris.csv", row.names = F)
```

```
...
```

## 1. Distribució dels estudiants i dels joves és igual?

```
```{r}
```

```
joves <- readxl::read_xlsx("~/Desktop/TFG/Dades/pob_barris_2016.xlsx")
```

```
dadescomp <- data.frame(barris$Num_barri[-c(70,71)], barris$Total_act[-c(70,71)],  
joves$joves[-74], barris$Districte[-c(70,71)])
```

```
dadescomp[69,2] <- dadescomp[69,2] + barris$Total_act[70]
```

```
colnames(dadescomp) <- c("num_barri", "num_estudiants", "num_joves", "districte")
```

```
# Taula
```

```
suma_est <- sum(dadescomp$num_estudiants)
```

```
suma_bar <- sum(dadescomp$num_joves)
```

```
prop_est <- c()
```

```
prop_bar <- c()
```

```
for(i in 1:nrow(dadescomp)){
```

```
  prop_est[i] <- dadescomp$num_estudiants[i]/suma_est
```

```
  prop_bar[i] <- dadescomp$num_joves[i]/suma_bar
```

```
}
```

```
prop_est_p <- prop_est*100
```

```
prop_bar_p <- prop_bar*100
```

```
dadescomp <- cbind(dadescomp, round(prop_est_p, digits = 2), round(prop_bar_p,  
digits = 2))
```

```
colnames(dadescomp) <- c("num_barri", "num_estudiants", "num_joves", "districte",  
"proporcio_est", "proporcio_joves")
```

```
write.csv2(dadescomp, "proporcions_estudiants.csv", row.names = F)
```

```
...
```

## 2. Dependència entre % d'estudiants per barri i renda:

```
```{r}
```

```
dadescomp2 <- data.frame(barris$Num_barri, barris$Total_act, barris$Districte)
```

```
colnames(dadescomp2) <- c("num_barri", "num_estudiants", "districte")
```

```

# Taula
suma_est <- sum(dadescomp$num_estudiants)
prop_est <- c()

for(i in 1:nrow(dadescomp2)){
  prop_est[i] <- dadescomp2$num_estudiants[i]/suma_est
}
prop_est_p <- prop_est*100

rendab <- readxl::read_xlsx("~/Desktop/TFG/Dades/Renda - barris.xlsx")
dep_renda <- data.frame(prop_est_p, "renda" = rendab$renda16)

# Guardar taula amb percentatges
write.csv2(dep_renda, "Percentatges estudiants per barri.csv")

# Gràfic linealitat
ggplot(data=dep_renda, aes(x=renda, y=prop_est_p)) + geom_point()

# Test
cor.test(dep_renda$renda, dep_renda$prop_est_p, method = "spearman")
...

```

### 3. Anàlisi d'independència entre factors branca - barri:

```

```{r}
library(vcd)

taula <- data.frame()
aux <- as.numeric(c(1:70,71,72,73,74,75))
for(i in 1:75){
  k <- aux[i]
  sub <- subset(barris_prop, barris == k)
  taula <- rbind(taula, sub$recompte)
}

```

```
dimnames(taula) = list(Barri = c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10", "11", "12",
"13", "14", "15", "16", "17", "18", "19", "20", "21", "22", "23", "24", "25", "26", "27",
"28", "29", "30", "31", "32", "33", "34", "35", "36", "37", "38", "39", "40", "41", "42",
"43", "44", "45", "46", "47", "48", "49", "50", "51", "52", "53", "54", "55", "56", "57",
"58", "59", "60", "61", "62", "63", "64", "65", "66", "67", "68", "69", "71", "72", "73",
"74", "75"), Branca = c("101", "102", "103", "201", "202", "203", "204", "205", "301",
"302", "401", "402", "403", "501", "502", "503", "504", "999"))
```

```
chisq.test(x = taula)
```

```
chisq.test(taula, simulate.p.value = TRUE, B = 5000)
```

```
# chisq.test(x = taula)$residuals # analisi residus de pearson
```

```
# chisq.test(x = taula)$stdres # els que s'allunyen més de zero són els que donen
problemes
```

```
...
```

#### 4. Anàlisi d'independència entre factors branca - sexe:

```
``{r}
```

```
# Gràfic
```

```
library(ggplot2)
```

```
t <- as.data.frame(table(dd$P1N, by = dd$sexe))
```

```
levels(t$by) <- c("Homes", "Dones")
```

```
levels(t$Var1) <- c("Filosofia i història", "Llengües i literatures", "Arts i disseny",
"Economia, Empresa i turisme", "Dret, laboral i polítiques", "Comunicació i
documentació", "Educació", "Intervenció social", "Ciències biològiques i de la terra",
"Ciències experimentals i matemàtiques", "Infermeria i salut", "Psicologia i teràpia",
"Medicina i ciències biomèdiques", "Arquitectura, construcció i civil", "Tecnologies
industrials", "TIC", "Agrícola, forestal i pesca", "Grau obert")
```

```
ggplot(t, aes(Var1, Freq)) + geom_bar(stat = "identity", fill = "#00538C") + labs(title
="Barplot gènere segons branca d'estudi", x = "Branca d'estudi", y = "Freqüència") +
theme_minimal() + coord_flip() + facet_grid(~by)
```

```
# Test
```

```
t2 <- data.frame(t$Freq[1:18], t$Freq[19:36])
```

```
dimnames(t2) = list( Branca = c("101", "102", "103", "201", "202", "203", "204", "205",
"301", "302", "401", "402", "403", "501", "502", "503", "504", "999"), Sexe = c("Home",
"Dona"))
```



```
chisq.test(x = t2)
```

```
...
```

## 5. Anàlisi gràfic de branca - nacionalita (ESP-NO ESP):

```
```{r}
```

```
dd2 <- dd
```

```
for(i in 1:nrow(dd2)){
```

```
  if(dd2$nac[i] != 1)
```

```
    {dd2$nac[i] <- 0}
```

```
}
```

```
t3 <- as.data.frame(table(dd2$P1N, by = dd2$nac))
```

```
tot_esp <- sum(subset(t3, by == 1)$Freq)
```

```
tot_no_esp <- sum(subset(t3, by == 0)$Freq)
```

```
for(i in 1:18){
```

```
  t3$Freq[i] <- t3$Freq[i]/tot_no_esp*100
```

```
}
```

```
for(i in 19:36){
```

```
  t3$Freq[i] <- t3$Freq[i]/tot_esp*100
```

```
}
```

```
levels(t3$by) <- c("Nacionalitat no espanyola", "Nacionalitat espanyola")
```

```
levels(t3$Var1) <- c("Filosofia i història", "Llengües i literatures", "Arts i disseny",  
"Economia, Empresa i turisme", "Dret, laboral i polítiques", "Comunicació i  
documentació", "Educació", "Intervenció social", "Ciències biològiques i de la terra",  
"Ciències experimentals i matemàtiques", "Infermeria i salut", "Psicologia i teràpia",  
"Medicina i ciències biomèdiques", "Arquitectura, construcció i civil", "Tecnologies  
industrials", "TIC", "Agrícola, forestal i pesca", "Grau obert")
```

```
ggplot(t3, aes(Var1, Freq)) + geom_bar(stat = "identity", fill = "#00538C") + labs(title  
="Barplot nacionalitat segons branca d'estudi", x = "Branca d'estudi", y = "Percentatge")  
+ theme_minimal() + coord_flip() + facet_grid(~by)
```

```
...
```

## 6. Anàlisi correlació branca - Índex de renda:

```
```{r}
```

```
rendab <- readxl::read_xlsx("~/Desktop/TFG/Dades/Renda - barris.xlsx")
```

```
for(i in 1:nrow(barris_prop)){
```

```

for(j in 1:nrow(rendab)){
  if(barris_prop$barris[i] == rendab$Num_Barri[j]){
    barris_prop[i,7] <- rendab[j,3]
  }
  for(k in 1:nrow(barris)){
    if(barris_prop$barris[i] == barris$Num_barri[k]){
      barris_prop[i,8] <- barris$Districte[k]
    }
  }
}

for(i in 1:nrow(barris_prop)){
  for(j in 1:nrow(distr)){
    if(barris_prop$V8[i] == distr$num[j]){
      barris_prop[i,9] <- distr$districtes[j]
    }
  }
}

colnames(barris_prop) <- c("branca", "recompte", "barris", "nom_barri", "proporcio",
"percentatge", "renda16", "num_distr", "Districte")

library(ggplot2)
library(gridExtra)

d101 <- subset(barris_prop, branca == "101")
gd101 <- ggplot(data = d101, aes(x = renda16, y = percentatge)) + geom_point(aes(color
= Districte)) + labs(title = "Branca: Filosofia i història", x = "Índex de renda", y =
"Percentatge")
cor1 <- cor.test(d101$proporcio, d101$renda16, method = "spearman")

d102 <- subset(barris_prop, branca == "102")

```

```
gd102 <- ggplot(data = d102, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Llengües i literatures", x = "Índex de renda", y = "Percentatge")
```

```
cor2 <- cor.test(d102$proporcio, d102$renda16, method = "spearman")
```

```
d103 <- subset(barris_prop, branca == "103")
```

```
gd103 <- ggplot(data = d103, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Arts i disseny", x = "Índex de renda", y = "Percentatge")
```

```
cor3 <- cor.test(d103$proporcio, d103$renda16, method = "spearman")
```

```
d201 <- subset(barris_prop, branca == "201")
```

```
gd201 <- ggplot(data = d201, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Economia, empresa i turisme", x = "Índex de renda", y = "Percentatge")
```

```
cor4 <- cor.test(d201$proporcio, d201$renda16, method = "spearman")
```

```
d202 <- subset(barris_prop, branca == "202")
```

```
gd202 <- ggplot(data = d202, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Dret, laboral i polítiques", x = "Índex de renda", y = "Percentatge")
```

```
cor5 <- cor.test(d202$proporcio, d202$renda16, method = "spearman")
```

```
d203 <- subset(barris_prop, branca == "203")
```

```
gd203 <- ggplot(data = d203, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Comunicació i documentació", x = "Índex de renda", y = "Percentatge")
```

```
cor6 <- cor.test(d203$proporcio, d203$renda16, method = "spearman")
```

```
d204 <- subset(barris_prop, branca == "204")
```

```
gd204 <- ggplot(data = d204, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Educació", x = "Índex de renda", y = "Percentatge")
```

```
cor7 <- cor.test(d204$proporcio, d204$renda16, method = "spearman")
```

```
d205 <- subset(barris_prop, branca == "205")
```

```
gd205 <- ggplot(data = d205, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Intervenció social", x = "Índex de renda", y = "Percentatge")
```

```

cor8 <- cor.test(d205$proporcio, d205$renda16, method = "spearman")

d301 <- subset(barris_prop, branca == "301")
gd301 <- ggplot(data = d301, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Ciències biològiques i de la terra", x = "Índex de renda", y = "Percentatge")
cor9 <- cor.test(d301$proporcio, d301$renda16, method = "spearman")

d302 <- subset(barris_prop, branca == "302")
gd302 <- ggplot(data = d302, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Ciències experimentals i matemàtiques", x = "Índex de renda", y = "Percentatge")
cor10 <- cor.test(d302$proporcio, d302$renda16, method = "spearman")

d401 <- subset(barris_prop, branca == "401")
gd401 <- ggplot(data = d401, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Infermeria i salut", x = "Índex de renda", y = "Percentatge")
cor11 <- cor.test(d401$proporcio, d401$renda16, method = "spearman")

d402 <- subset(barris_prop, branca == "402")
gd402 <- ggplot(data = d402, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Psicologia i teràpia", x = "Índex de renda", y = "Percentatge")
cor12 <- cor.test(d402$proporcio, d402$renda16, method = "spearman")

d403 <- subset(barris_prop, branca == "403")
gd403 <- ggplot(data = d403, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Medicina i ciències biomèdiques", x = "Índex de renda", y = "Percentatge")
cor13 <- cor.test(d403$proporcio, d403$renda16, method = "spearman")

d501 <- subset(barris_prop, branca == "501")
gd501 <- ggplot(data = d501, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Arquitectura, construcció i civil", x = "Índex de renda", y = "Percentatge")
cor14 <- cor.test(d501$proporcio, d501$renda16, method = "spearman")

```

```

d502 <- subset(barris_prop, branca == "502")
gd502 <- ggplot(data = d502, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Tecnologies industrials", x = "Índex de renda", y = "Percentatge")
cor15 <- cor.test(d502$proporcio, d502$renda16, method = "spearman")

d503 <- subset(barris_prop, branca == "503")
gd503 <- ggplot(data = d503, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: TIC", x = "Índex de renda", y = "Percentatge")
cor16 <- cor.test(d503$proporcio, d503$renda16, method = "spearman")

d504 <- subset(barris_prop, branca == "504")
gd504 <- ggplot(data = d504, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Agrícola, forestal i pesca", x = "Índex de renda", y = "Percentatge")
cor17 <- cor.test(d504$proporcio, d504$renda16, method = "spearman")

d999 <- subset(barris_prop, branca == "999")
gd999 <- ggplot(data = d999, aes(x = renda16, y = percentatge)) + geom_point(aes(color = Districte)) + labs(title = "Branca: Grau obert", x = "Índex de renda", y = "Percentatge")
cor18 <- cor.test(d999$proporcio, d999$renda16, method = "spearman")

grid.arrange(gd101, gd102, nrow=2)
grid.arrange(gd103, gd201, nrow=2)
grid.arrange(gd202, gd203, nrow=2)
grid.arrange(gd204, gd205, nrow=2)
grid.arrange(gd301, gd302, nrow=2)
grid.arrange(gd401, gd402, nrow=2)
grid.arrange(gd403, gd501, nrow=2)
grid.arrange(gd502, gd503, nrow=2)
grid.arrange(gd504, gd999, nrow=2)

llistat_cor <- c(cor1$estimate[1][[1]], cor2$estimate[1][[1]], cor3$estimate[1][[1]],
cor4$estimate[1][[1]], cor5$estimate[1][[1]], cor6$estimate[1][[1]],
cor7$estimate[1][[1]], cor8$estimate[1][[1]], cor9$estimate[1][[1]],

```

```
cor10$estimate[1][[1]],      cor11$estimate[1][[1]],      cor12$estimate[1][[1]],
cor13$estimate[1][[1]],      cor14$estimate[1][[1]],      cor15$estimate[1][[1]],
cor16$estimate[1][[1]], cor17$estimate[1][[1]], cor18$estimate[1][[1]])
```

```
llistat_cor <- as.data.frame(cbind(c(1:18), llistat_cor))
```

```
colnames(llistat_cor) <- c("X", "Coeficient_cor")
```

```
ggplot(llistat_cor, aes(X, Coeficient_cor)) + geom_point(col = "#00538C") +
theme_minimal() + geom_hline(yintercept = c(0.22, -0.22), lty = 2, col = "grey") +
labs(title = "Coeficients de correlació de Spearman", x = "", y = "Coeficient")
```

```
...
```

## 7. Relacionar l'edat amb el % d'estudiants

```
```{r}
```

```
dd3 <- dd
```

```
for(i in 1:nrow(dd3)){
```

```
  if(dd3$edatT[i] <=18){
```

```
    dd3$edatT[i] <- 1
```

```
  } else if(dd3$edatT[i] >=19 && dd3$edatT[i] <=25){
```

```
    dd3$edatT[i] <- 2
```

```
  } else if(dd3$edatT[i] > 25){
```

```
    dd3$edatT[i] <- 3
```

```
  }
```

```
}
```

```
t3 <- as.data.frame(table(dd3$P1N, by = dd3$edatT))
```

```
levels(t3$by) <- c("[16,18]", "[19-25]", "[26,74]")
```

```
levels(t3$Var1) <- c("Filosofia i història", "Llengües i literatures", "Arts i disseny",
"Economia, Empresa i turisme", "Dret, laboral i polítiques", "Comunicació i
documentació", "Educació", "Intervenció social", "Ciències biològiques i de la terra",
"Ciències experimentals i matemàtiques", "Infermeria i salut", "Psicologia i teràpia",
"Medicina i ciències biomèdiques", "Arquitectura, construcció i civil", "Tecnologies
industrials", "TIC", "Agrícola, forestal i pesca", "Grau obert")
```

```
ggplot(t3, aes(Var1, Freq)) + geom_bar(stat = "identity", fill = "#00538C") + labs(title =
"Barplot grups d'edat segons branca d'estudi", x = "Branca d'estudi", y = "Freqüència")
+ theme_minimal() + coord_flip() + facet_grid(~by)
```

```
# Test  
t4 <- data.frame(t3$Freq[1:18], t3$Freq[19:36], t3$Freq[37:54])  
  
dimnames(t4) = list( Branca = c("101", "102", "103", "201", "202", "203", "204", "205",  
"301", "302", "401", "402", "403", "501", "502", "503", "504", "999"), Edat = c("1", "2",  
"3"))  
  
chisq.test(x = t4)  
...
```

